

# EXECUTIVE COACHING OUTCOME RESEARCH IN A FIELD SETTING: A NEAR-RANDOMIZED CONTROLLED TRIAL STUDY IN A GLOBAL HEALTHCARE CORPORATION

ERIK DE HAAN

Ashridge/Hult International Business School, Berkhamsted (Herts.), United Kingdom  
VU University Amsterdam, The Netherlands

DAVID E. GRAY

University of Greenwich, London, United Kingdom

SALLY BONNEYWELL

Bonneywell Development Ltd., London, United Kingdom

Executive coaching is widely used as a tool to develop leaders and organizations. However, despite the popularity of coaching, the effect of coaching on relevant leadership performance measures remains unclear. We report on the development of a model for coaching effectiveness predicting that aspects common to all approaches are the main active ingredients, mediated by the working alliance as the single-best predicting common aspect. To test this model, we conducted a (waiting-list) randomized controlled trial within a global healthcare corporation, involving 180 coachees, 66 coaches, and 140 line managers of the coachees, which we contend is the largest randomized controlled trial in executive coaching to date. The study demonstrated substantial impact with effect sizes up to  $d = 1.1$ . We also tested several aspects that are common to coaching contracts and found more evidence for the central importance of the strength of the working alliance as seen from both the coachee and (to a lesser extent) the coach perspectives, which mediated the effects for coachee well-being and perceived social support. In addition, we found some significant impact for coachee self-efficacy, resilience and “bright side” personality aspects. Finally, we found some first significant indications that personality and career derailment aspects may show demonstrable improvement through coaching. We discuss the implications of these findings for coaching theory and practice. Overall, our study demonstrates the effectiveness of coaching in an organizational setting, the importance of the working alliance, and that coaching may mitigate the risks of leaders derailing in large organizations.

The use of executive coaching has increased substantially in organizations over the last 20 years, such that coaching is now firmly established as a development opportunity for senior leaders in large organizations (Ridler, 2016). One global study of coaching estimates there are over 50,000 coach practitioners worldwide, with a further 10,000 managers taking their internal coaching role as far as becoming a formal member of a coach association

(ICF, 2016). Moreover, executive coaching is now managed and adopted independently from adjacent fields such as organization-development consultancy and leadership-development programs, as well as integrated within these. Over these 2 decades, increasingly rigorous standards and qualifications have been developed, and international coaching associations and research centers have matured. However, as Armstrong (2011) points out, indications suggest that in many parts of the world coaching has not gone beyond the introduction or growth phase of the product life cycle. Nevertheless, coaching presents a unique challenge to the field of management research because it represents a new practice in organizations (Kenworthy, Passarelli & Van Oosten, 2014).

---

We would like to thank Hogan Inc., and Brandon Ferrell in particular, for generous help in use and administration of three psychometric instruments (HPI, HDS, MVPI). We also gratefully acknowledge Audrey Vandeborne for data collection, Sue Gammons for data management, and Viktor Nilsson for data analysis.

A wide variety of research into coaching processes and outcomes has now been undertaken (for recent overviews, see Ely et al. 2010; De Haan & Duckworth, 2013; Blackman, Morsardo, & Gray, 2016; Grover & Furnham, 2016; and Athanasopoulou & Dopson, 2018). Theeboom, Beersma, and van Vianen (2014) and Jones, Woods, and Guillaume (2015) have even attempted some early meta-analyses. However, as Liu and Batt (2010) point out, much of the research into executive coaching has used managers on Leadership or MBA courses as subjects, as opposed to experienced managers in an organizational setting (as in the current study). Additionally, most studies so far have investigated links between coaching and performance and other important criteria by utilizing cross-sectional designs and subjective self-report measures that may obscure the actual benefits of coaching (Dahling, Taylor, Chau, & Dwight, 2016). Furthermore, research has not yet reached the same rigor and status as in other helping professions, as shown by the fact that the application of the “gold standard in clinical research,” the randomized controlled trial (Passmore & Theeboom, 2015) in coaching research, occurs only in a handful of small-scale studies, as will be shown below in the Literature Review section. Indeed, very few corporate coaching programs have even been formally evaluated (McDermott, Levenson, & Newton, 2007).

Despite the limited harvest so far, we argue that it is important to undertake large-scale randomized controlled trials (RCTs) in executive and workplace coaching. In the first place, RCTs are the standard means to collect evidence of an intervention that can be attributed to that intervention itself, and therefore, are the best way to check if the increasing corporate expenditure in coaching is warranted. Secondly, RCTs can provide clues as to the active ingredients (the independent variables that significantly interact with coaching outcomes) in coaching and can show how such ingredients interact to moderate or mediate the effects of coaching, thus helping maximize investment yields in coaching. In other words, RCT designs can help to find out which individual coachee might benefit most from the intervention, or which preparation, or way of working, is optimal for coaches. Evidence of this kind can help to improve the coaching intervention to enhance those techniques or types of contracts that give the highest return on investment. Many organizations are forced to be selective in terms of both the managers that may apply for workplace coaching and the (internal or external) coaches who

can provide this service. The results of RCTs can play an important role in making such delicate choices about investment in coaching and the matching of coachees and coaches. For example, on the grounds of outcome research to date, Page and De Haan (2014) suggest that it is better not to undertake matching based on demographics or personality fit as most of the evidence is to the contrary; namely, that self-selection based on chemistry sessions by the prospective coachees themselves works best. Third, RCTs can look into the uniqueness of the executive-coaching intervention. Coaching is a tailored intervention for individual leaders, and as such it is hoped that coaching can make a difference in important and highly personal aspects of their leadership, such as the leader's resilience, responsibility-taking, or ethical use of power. As has been suggested before (Levenson, 2009; Nelson & Hogan, 2009), coaching might be most useful to “challenge” leaders and help preclude the breakdown of resilience, responsibility, or leadership ethics. In other words, coaching might be uniquely targeted at reducing “personality derailment tendencies” of executives, which can be very costly to large organizations (Kaiser, LeBreton, & Hogan, 2015), when tested by an RCT study with the right choice of dependent variables.

We report for the first time on an RCT study in an industry-based setting, within a large global corporation that manages more than 1400 coaching contracts on a yearly basis. This study contributes to the existing coaching outcome literature in three important ways: First, it provides the largest scale randomized-controlled trial to date in the literature; second, it studies the outcome of executive (group) coaching in a naturalistic corporate setting while still studying coaching uncontaminated by other leadership development activities; and third, is the first to model and demonstrate the impact of coaching on leadership-derailment patterns, by measuring these personality characteristics before and after the intervention.

## LITERATURE REVIEW

### Overview of Coaching Research With Control Groups

As in the current research, a number of studies have adopted a quasi-experimental design (Smither, London, Flautt, Vargas, & Kucine, 2003; Evers, Brouwers, & Tomic, 2006; Nieminen, Smerek, Kotrba, & Denison, 2013; MacKie, 2014). Smither

et al. (2003) is still one of the most thorough studies on the impact of executive coaching. Their design involved a (nonrandomized) control group, and conclusions were based on more objective criteria than evaluations by the coachees; namely, use of evaluations by independent researchers together with coachees' superiors, colleagues, and staff (multisource feedback). This research involved 1202 senior managers in one multinational organization with two consecutive years of 360-degree feedback. The researchers found that managers who worked with an executive coach were significantly more likely than other managers to (1) set specific goals (Cohen's  $d = 0.16$ ;  $p < 0.01$ ); (2) solicit ideas for improvements from their superiors ( $d = 0.36$ ;  $p < 0.01$ ); and (3) obtain higher ratings from direct reports and superiors in the second year ( $d = 0.17$ ;  $p < 0.05$ ). This was a significant result particularly given there were no more than "two or three" coaching sessions per coachee (Smither et al., 2003: 29).

Evers et al. (2006) measured self-efficacy beliefs and outcome expectancies on each of three dimensions. Their study compared a pre- and post-intervention measurement and also involved a (nonrandomized) control group. Although the sample was quite small (30 managers in both the experimental and the control group), they did find some empirical evidence for a positive outcome of the coaching intervention. There was a significant increment for the coached group over the control group for one of the three dimensions in both self-efficacy beliefs ("setting one's own goals") and outcome expectancies ("acting in a balanced way") ( $d \approx 0.5$  with  $p < 0.05$ ). However, the intervention was short with an average of only four coaching sessions.

In another small-scale quasi-experimental study, MacKie (2014) assigned executives and senior managers in a not-for-profit organization nonrandomly to either a coaching or a waiting-list cohort, with cohort sizes of 14 and 17, respectively. Both groups received six sessions of strengths-based coaching with experienced external coaches. The results revealed that participants experienced highly statistically significant increases in their transformational leadership behavior after coaching (Cohen's  $d = 0.9$ ;  $p < 0.01$ ), and this difference was reported at all levels within the organization although not by the participants themselves.

Nieminen et al.'s (2013) quasi-experimental study followed 469 managers, 227 of whom received four or five sessions of executive coaching and a multisource feedback (MSF) session, while 242

received only the MSF. Results indicated that managers in both groups improved similarly as rated by direct reports, peers, and supervisors, while only those managers who received the executive coaching intervention improved according to self-ratings (Cohen's  $d = 0.21$ ;  $p < 0.01$ ). Although managers had been allocated to the two groups based on non-random selection methods, those in the coached group were comprised of managers who had recently been promoted and had fewer years of leadership than the noncoached group. Similarly, Ladegard and Gjerde (2014) studied leaders who had received coaching for eight sessions by experienced external coaches, although the numbers were limited, with only 18 coachees and six nonrandomized controls (80 of their direct reports were studied for subordinate feedback). There was a statistically significant relationship between managers' trust in subordinates and reduced turnover intentions of the subordinates. Bozer, Sarros, and Santora (2013) studied coachee characteristics in a quasi-experimental field study with 12 coaching sessions for 72 coachees and 29 of their peers in a non-randomized control group, with data from their 28 direct supervisors. They found clear effectiveness in terms of career satisfaction (Cohen's  $d = 0.87$ ;  $p < 0.001$ ), but surprisingly the control group performed better in terms of supervisor-rated task performance (Cohen's  $d = 0.54$ ;  $p < 0.05$ ). Bozer et al. (2013) also found a positive relationship of learning goal orientation, self-efficacy, and pre-coaching motivation to coaching effectiveness. Moen and Skaalvik (2009) trained 11 senior leaders through a coaching program, group coaching and external executive coaching, who then formed the experimental group, to coach their 52 middle managers, while in a nonrandomized control group eight executives worked with 56 middle managers without further training. They show significant findings in terms of managers' self-efficacy, goal-setting behaviors, intrapersonal causal attributions of success, and need satisfaction.

Following on from a PhD project with three small randomized controlled trials (groups of around 20) that demonstrated some significant impact of a coaching workshop and five follow-up coaching sessions on objective academic results and anxiety self-ratings (Grant, 2003), a series of truly randomized but small-scale controlled trials was undertaken in health, educational, and life-coaching settings. Of these, Green, Oades, and Grant (2006), Green, Grant, and Rynsaardt (2007), and Spence and Grant (2007) employed randomized controlled trials with waiting

lists for  $N = 58$ ,  $N = 56$ , and  $N = 63$  participants. Note, however, that in the first two studies,  $N = 6$  and 7 participants withdrew before the second questionnaire, and in the third study there were two treatment groups, so again numbers in every cohort are fewer than 30. Each study used only self-report measures, introducing the risk of same-source biases. They found some significant effects, with the largest found for group coaching (Green et al., 2006). Similarly, Grant, Curtayne, and Burton (2009) studied  $N = 41$  healthcare managers, with approximately half assigned to a random waiting-list control group and the other half receiving executive coaching. They found that coaching significantly enhanced goal attainment, resilience, workplace well-being, and reduced depression and stress (all self-scored; all  $p < 0.05$ ). Finally, Grant, Green, and Rynsaardt (2010) studied  $N = 44$  high-school teachers, of which approximately half were assigned to a random waiting-list control group and the other half received work-related coaching. They found that coaching again significantly enhanced self-scored goal attainment, resilience, workplace well-being, and reduced depression and stress ( $d$  varying between 0.4 and 1.5). Pre- and post- *self-ratings* in leadership styles were also significantly different (with  $d$ 's around 0.5); however, these rating differences remained insignificant in *peer ratings*. So, in the only article in this series where self- and other-ratings are compared (Grant et al., 2010), the results for other-ratings are insignificant.

Duijts, Kant, Van den Brandt, and Swaen (2008) found similar results in a randomized controlled investigation into the effectiveness of seven-to-nine sessions of "preventive" work-related coaching in terms of reducing sickness absence due to psychosocial health complaints.  $N = 151$  employees (all at risk for sickness absence) found themselves randomly assigned to two almost equal-size intervention and no-intervention control groups, although only 37 of them participated fully in the coaching intervention. Again, the intervention group self-reported improved health, while the findings on most objective measures were not significant. However, statistically significant reductions in sickness absence could indeed be demonstrated in this group (2.5 days less than in the intervention group off work on average in the year following coaching;  $p < 0.01$ ).

Finally, two randomized controlled trials in driving instruction are worth noting, where a "coaching" approach was compared with a traditional instruction-only approach (Passmore & Velez, 2012,

with 327 drivers, and Passmore & Rehman, 2012, with 208 drivers). These studies show some statistically significant improvements (in terms of number of hours driving that pupils needed to pass the test, the number of tests taken by pupils before passing, and the likelihood of pupils to pass on the first attempt; Passmore & Rehman, 2012), when instructors adopted coaching-type interventions. However, there were no significant findings in the case of one time only coaching sessions (Passmore & Velez, 2012).

### Limitations of Past Research

#### ***Most earlier studies contain major weaknesses including:***

- (1) The use of nonrandomized samples (e.g., Smither et al., 2003; Nieminen et al., 2013; Ladegard & Gjerde, 2014; MacKie, 2014; Evers et al., 2006);
- (2) The scarcity of reliable data because of small sample sizes (e.g., Bozer et al., 2013; Moen & Skaalvik, 2009; MacKie, 2014; Evers et al., 2006);
- (3) The predominance of self-score measures, inviting the possibility of same-source bias (e.g., while the Duijts et al., 2008, and Grant et al., 2010, studies used randomized allocation, we have noted that only the self-reported improvements were statistically significant, the results on objective measures were not significant).

Furthermore, O'Neil, Hopkins, and Bilimoria (2015) comment that few studies have explored the impact of executive coaching on female leaders, as women remain under-represented in leadership roles. Exceptions are the studies by Gray and Goregaokar (2010), Bozer, Baek-Kyoo, and Santora (2015), and De Haan, Grant, Burger, and Eriksson (2016), which analyze executive coaching in relation to gender. However, these studies all conclude that the coach-coachee match has little effect on coaching outcomes.

We therefore argue that what is needed to really get to the heart of coaching effectiveness (CE) are large-scale studies where membership of target and control groups are randomly assigned, and which make use of a variety of outcome measures and not just self-scoring. We offer our study as one such design.

### HYPOTHESIS DEVELOPMENT

Our study involved two consecutive groups on a leadership-development program ("Accelerating Difference," see Methodology) designed to increase

the ratio of female leaders at all leadership levels. The two starting points of April and September provided ideal conditions for a “waiting list control group” study design. The study examines the relative impact of various “common factors” (i.e., aspects common to all coaching approaches, see De Haan, 2008), with the help of the largest randomized controlled trial to date: 89 female leaders in the target group at T2, and 72 female leaders in the control group, with at least 114 corresponding questionnaires from their 66 coaches and 115 from their 140 line managers. By asking coachees, coaches, and line managers to rate outcomes on the same scales, we have three independent measures of coaching outcomes. We check for the fact that some of the data structure is multilevel, although average nesting is well under three for the coaches (i.e., a coach does not work with more than three coachees on average) and well under two for the managers. We emphasize that the fact that our sample includes only female leaders is both unique and keeps gender constant for comparability purposes. We check for the influence of gender by comparing female–male and female–female coaching relationships. We discuss these and other potential drawbacks of this sample design later under Limitations.

### Demonstrating Coaching Effectiveness

Based on the effect sizes found in the studies summarized above, we expect coaching to be demonstrably effective with a considerably smaller effect size than in psychotherapy, where  $d$  is around 0.85 (see De Haan, 2008, for a summary).

*H1: Coaching Effectiveness (CE) scores are significantly different with outcome measurements' effect sizes below 0.85 between coachee and control groups, (a) as measured by the coachee, (b) as measured by the coach, and (c) as measured by the line manager.*

The most consistent active ingredient studied in earlier coaching-outcome research has been the strength of the coach–coachee relationship, operationalized as coach–coachee chemistry (Hooijberg & Lane, 2009) or working alliance (see e.g., Baron & Morin, 2009; De Haan et al., 2013, 2016). We expect working alliance (WAI) as a measure of the quality of the coaching relationship to be the best predictor of present and future effectiveness (see Horvath & Symonds, 1991, for a summary of how WAI has been shown to be a powerful active ingredient in psychotherapy). Longitudinal analyses of the levels of WAI in both more and less successful

counseling and therapy (e.g., Horvath & Marx, 1990; Safran, Crocker, McMain, & Murray, 1990; Safran, Muran, & Wallner Samstag, 1993) provided evidence for the existence of relational rupture-repair cycles in successful counseling and therapy. Meta-analysis has shown (Martin, Garske, & Davis, 2000) that patients tend to find the working alliance more stable than therapists and observers, and it is generally found to be important to distinguish between the perspectives of coaches, coachees, or observers, as they tend to be nearly independent (De Haan, Bertie, Day, & Sills, 2010).

*H2: The strength of the coaching working alliance (WAI) will predict current and even future CE scores (as measured by coachee, coach and line manager), (a) when WAI is measured by the coachee, (b) but also when measured by the coach.*

General self-efficacy is the belief in one's competence to cope with a broad range of stressful or challenging demands (Luszczynska, Scholz, & Schwarzer, 2005). Self-efficacy has been found to be a valid construct in a broad range of ethnic and cultural settings globally (Luszczynska et al., 2005; Schwarzer, Bäßler, Kwiatek, Schröder, & Zhang, 1997). Thus, general self-efficacy may be an important variable that mediates the effect of the coach–coachee relationship on CE. Research by Anderson and Betz (2001) showed that a person's self-efficacy expectations have a direct bearing on his or her personal and career development. Sherer et al. (1982) found that personal self-efficacy expectations are often regarded as primary determinants of behavioral change. Within the coaching outcome literature, Stewart, Palmer, Wilkin, and Kerrin (2008) and De Haan et al. (2013, 2016) have shown that general self-efficacy predicts coaching effectiveness, while Evers et al. (2006) found that coaching enhanced self-efficacy, and Grant (2014) also demonstrated that executive coaching during a period of organizational change increased leadership self-efficacy. Hence, a coachee's general self-efficacy ratings can be expected to predict his or her coaching outcomes. Roehrl and Strouse (2008) found social support is also a statistically significant extra-therapeutic factor ( $d = 0.22$ ). Subjective or psychological well-being has been found to have a positive impact on workplace productivity (Bryson, Forth, & Stokes, 2014) and can now be measured with a well-validated short questionnaire (Tennant et al., 2007). Grant et al. (2009, 2010) showed that coaching can enhance (workplace) well-being and resilience.

*H3: Psychological well-being, resilience, social-support and self-efficacy of the coachee will predict CE as scored by coach, coachee, and line manager. These kinds of “preparedness” scores should also predict future CE scores.*

As to the behaviors of the coach, we expect these to look different from a coachee perspective and the coach’s perspective, yet we also expect moderate correlation between the two (De Haan & Nilsson, 2017). We think that common factors are much more relevant to outcome than specific technique, following many meta-analyses in the adjacent field of psychotherapy (summarized in De Haan, 2008), so we do not expect specific coaching behaviors to correlate with CE, either from the coach or coachee perspective. Similarly, with the help of a coachee-scored earlier version of the questionnaire that did not entail forced choice (De Haan, Culpin, & Curd, 2011), the coachee-scored coaching effectiveness was shown to correlate with all six coach behaviors to the same degree, confirming the hypothesis of common factors.

*H4: Coaching behaviors, as measured by both coachees and coaches, show only modest to zero predictions of coaching effectiveness.*

Personality potentially has an important role in the coach–coachee relationship, so we wanted to explore this issue here. Past research (Scoular & Linley, 2006) has found that personality dissimilarity as measured by (self-reported) MBTI (Myers-Briggs Type Inventory; Myers, McCaulley, Quenk, & Hammer, 1998) was related to CE. However, De Haan et al. (2013, 2016) did not find any significant correlations between personality differences or personality matching in terms of (self-reported) MBTI and CE. Stewart et al. (2008) also looked at how coachee personality factors correlate with (self-reported) coaching outcome, using the so-called “Big-5” personality factors (Digman, 1990) for 110 coachees and correlating these with self-reported coaching effectiveness. They found moderate positive and significant effects for Conscientiousness, Openness, and Emotional Stability. In other words, in the Stewart et al. (2008) study, coaching appeared to be most effective for executives scoring high in Emotional Stability, Conscientiousness, and Openness to experience, and Conscientiousness was also a predictor of the coachees’ use of coaching outcomes in their day-to-day managerial activities.

*H5: The impact of personality aspects will be mixed and small:*

*MBTI (both coach–coachee personality dissimilarity and matching) will not predict CE;*

*Neither will a person’s values or personality derailment tendencies (as measured by Hogan’s MVPI and HDS);*

*However, because of the link with the earlier findings of the Big-5 personality factors, there will be a small significant effect coming from some “bright side” personality characteristics described by Hogan’s HPI.*

There has long been an expectation that executive coaching can be particularly helpful with understanding and changing unhelpful, maladapted, and hidden personality characteristics of executives (see e.g., Kets de Vries, 2006). Because of the personalized nature, the high confidentiality, and possibility for deep understanding and challenge, coaching seems to work at relational and personal depths, touching on personality derailment patterns of executives, as argued by Nelson and Hogan (2009). For this reason, we expect executive coaching to have a differential effect on personality derailment tendencies. This would support the use of executive coaching for personalized leadership development and for mitigating risks to the organization stemming from such proven derailment factors (Hogan & Kaiser, 2005).

*H6: Coaching has a demonstrably positive effect on some personality derailment tendencies, over and above the take-retake tendencies which can be measured in both experimental and control groups.*

## METHODOLOGY

### Research Context

The organization in which the study was situated is a global company, comprising approximately 100,000 employees based in over 120 countries. The company operates in the healthcare industry, specifically involving pharmaceuticals, vaccines, and “over the counter” healthcare products. Allocation to the two cohorts of the 14-month leadership-development program was essentially random, allowing us to take two consecutive cohorts for our randomized study design. We conducted measurements at three time points: before the start of both programs, before the start of the Cohort 2, and after the end of Cohort 1. The program consisted of individual executive coaching together with group

coaching, so we could study the effectiveness of coaching uncontaminated with any programmatic, educational inputs.

## Participants

**Coachee Details.** Line managers and HR leaders recruited participants locally, based on the following criteria:

- (1) *Future potential*: the participant had the potential to progress as a leader at least two levels within the organization;
- (2) *Ambition*: the participant actively wanted to progress her career and to take advantage of the development opportunities offered to her;
- (3) *Personal circumstances*: the participant had personal circumstances that enabled her to participate in development opportunities and to progress in the organization.

One hundred eighty of 209 coachees (80%) accepted the invitation to the research by completing the first research questionnaire. They were randomly assigned to one of two cohorts, so that we could create a 6-month time-lagged control group. One cohort commenced coaching in early April 2016, while the other commenced in late September, thereby ensuring an elapsed time of 6 months between measurements at T1, T2, and T3.

**Coach Details.** The coaches involved in the individual coaching were qualified internal coaches of three levels (having 3 to 20 days of formal training). The majority were line managers who have a paid “day job” within the organization, six of the internal coaches are qualified through a rigorous external training and assessment program, and they have regular externally facilitated group supervision and continuing professional development. The coachees were assigned one individual coach based on the seniority of the coachee; more senior-level coachees were randomly assigned more senior coaches, taking geographic location into account to maximize face-to-face coaching where possible.

The program’s design entailed each female leader receiving a maximum of 12 individual 1-hour coaching sessions (12 was deemed appropriate as the program lasted 14 months), which were conducted either in person, by Live Meeting (a web conferencing service via webcam), or by telephone. Due to the pressures of work and scheduling, in 10% of cases, coachees received 10 or 11 coaching sessions. In the same period, coachees also received six group-

coaching sessions (groups of 5–8), where group coaches were assigned in pairs so that each group had the benefit of having both an external and internal group coach. Attention was paid to the gender mix of the coaches. No groups had two male group coaches, 12 had one male and one female coach, and 18 had two female coaches. In approximately 20% of cases, there were only five sessions due to work or personal vacation commitments. Sixty-six coaches (response rate 61%) completed at least one questionnaire; 40 were female (61%), and 26 were male (39%).

**Line Manager Details.** All coachees had a line manager who was invited to participate in the research. Line managers attended an introductory webinar to understand their role and the purpose of Accelerating Difference. Line managers were expected to attend three formal meetings with their female participants and their coach. The first meeting also had the HR representative present. It took place between 1 and 2 months after the start of the coaching. The purpose was to agree on the coaching objectives for the individual coaching. The second meeting was with the participant and their coach with the purpose of reviewing progress and to see how the line manager could support the participant in his or her development. The final meeting was with the line manager, coach, participant, and HR representative, the purpose of which was to compare the progress against the coaching objectives, review the overall development of the female leader, and put in place support for her development following the ending of the initiative. Line managers were also expected to hold regular one-to-one meetings throughout the duration of the program. One hundred forty of 209 line managers (67%) completed the research questionnaire at T1.

## Measures

The first email referring to online questionnaires was sent to the coachee, coach, and line-manager before the first coaching conversation and was closed by the time coaches had been matched to coachees. The second questionnaires went out after 6 months when Cohort 2 was about to be matched to their coaches; similarly, this questionnaire was closed before the control group’s sessions began. The third measurement was again 6 months later, so it took place when the experimental group’s coaching had ended. Completions at T1 were between 18 March and 19 May 2016, at T2 between 3 October and 14 November 2016 and at T3 between 24 April and 30 May 2017.

**TABLE 1**  
**Measurement Table Showing Psychometrics Taken by Cohort (Coachee, Coach, Line Manager), and Time (T1, T2, T3).**

Study Variables:	Coachee T1	Coachee T2 + T3	Coach T1	Coach T2 + T3	Line manager T1	Line manager T2 + T3
Name	*		*		*	
Gender	*		*		*	
Nationality	*		*		*	
Country of residence	*		*		*	
MBTI Personality	*		*			
Name of Coachee/Coach	*	*	*	*	*	*
Outcome Expectations	*	*	*	*	*	*
Coaching Effectiveness (CE)	*	*		*	*	*
The Brief Resilience Scale (BRS)	*					
General Self-Efficacy (GSE)	*		*			
Psychological Well-Being Scale (WEMWBS)	*	*				
Perceived Social Support (PSS)	*	*				
Coaching Behaviors Questionnaire (CBQ)		*		*		
Working Alliance Inventory (WAI)		*		*		
Hogan "bright side" (HPI)	*	*				
Hogan "dark side" derailers (HDS)	*	*				
Hogan "motives, values, preferences" (MVPI)	*	*				

Note: Hogan questionnaires were only taken by the coachee at T1 and T2. Similarly, CBQ and WAI were taken at T2 and T3 for Cohort 1 and only at T3 for Cohort 2 (Control Group), because for the retake they first needed to experience coaching.

The content of each questionnaire per cohort is summarized in Table 1.

**Coaching Effectiveness (CE).** CE was assessed using four items on a 7-point response scale (range 0–7, *Strongly Disagree*, *Disagree*, *Slightly Disagree*, *Neither Agree nor Disagree*, *Slightly Agree*, *Agree*, *Strongly Agree*): “The outcome of my coaching objectives so far: (1) I have been successful in creating reflective space for me; (2) I have been successful in creating new insight for me; (3) Through (preparation for) coaching I have successfully engaged in new action or behavior; (4) I would consider this coaching journey successful.” We chose this scale because it has been used before and was shown to have good internal consistency (e.g., De Haan et al., 2013, 2016) and because it seems to map a wide range of areas of possible effectiveness of coaching. Moreover, this scale can be used *before* the coaching assignment commences, as a measure of effectiveness on these dimensions at the current time. Responses were calculated as the average score across these four items and Cronbach Alphas are shown in Table 2. We computed configural and metric invariance with GFI, CFI, NFI, and SRMR, which turned out to be excellent with all indices below 0.0001.

**Coach–Coachee Relationship.** The Working Alliance Inventory (WAI) was used as a measure of the strength of the coach–coachee relationship (Horvath & Greenberg, 1986). Prior permission was obtained to

adapt this 36-item instrument, which is used widely in therapy for measuring the strength and quality of the relationship between therapist and client, to measure the coach–coachee relationship. The WAI consist of three subscales: Task, Goal, and Bond, with 12 items each.

- The term *Task* refers to what coach and coachee agree need to be done in order for the coachee to reach his/her goals for coaching. A typical item is “I am clear as to what my coach wants me to do in these sessions.”
- The term *Goal* refers to the outcomes that the coach/coachee hopes to gain from coaching. A typical item is “The goals of these sessions are important to me.”
- The term *Bond* refers to the extent to which the coach/coachee trusts, respects, and feels confidence in the other person. A typical item is “I believe my coach is genuinely concerned for my welfare.”

**Perceived Social Support.** Social support was measured with help of Zimet’s Multidimensional Scale of Perceived Social Support (MSPSS; Zimet et al., 1988). The scale consists of 12 items on a 7-point Likert scale, mapping three subscales related to support from Family, Friends, and Significant Others, with sample items “my family really tries to help me,” “my friends really try to help me,” and



TABLE 2  
Scale Sizes, Means, Standard Deviations, Reliability Estimates (Cronbach Alphas  $\alpha$ , on the Diagonal), and Intercorrelations (Pearson's correlations) Among Study Variables

	1	2	3	4	5	6	7	8	9
1. CE (coach) #items = 4 M = 6.10 SD = 0.63	$\alpha = .91$								
2. CE (coach) #items = 4 M = 6.10 SD = 0.63	.05 (82)	$\alpha = .92$							
3. CE (line mgr.) #items = 4 M = 5.55 SD = 0.78	-.05 (95)	.17 (57)	$\alpha = .88$						
4. WAI (coach) #items = 36 M = 5.62 SD = 0.53	.11 (80)	.25* (97)	.32* (55)	$\alpha = .78$					
5. WAI (coach) #items = 36 M = 6.26 SD = 0.76	.68*** (80)	.12 (79)	.08 (89)	.14 (77)	$\alpha = .95$				
6. WEMWBS #items = 14 M = 55.50 SD = 6.25	.30*** (153)	-.01 (81)	.06 (94)	.13 (79)	.33*** (148)	$\alpha = .90$			
7. MSPSS #items = 12 M = 5.80 SD = 0.88	.31*** (151)	.01 (81)	.01 (92)	.06 (79)	.33*** (148)	.55*** (153)	$\alpha = .94$		
8. GSE #items = 10 M = 3.30 SD = 0.32	.04 (127)	.03 (79)	-.17 (102)	-.07 (77)	-.03 (120)	.06 (125)	.11 (124)	$\alpha = .79$	
9. BRS #items = 6 M = 22.93 SD = 3.36	.04 (127)	.12 (79)	-.03 (102)	.03 (77)	.02 (120)	.23** (125)	.28** (124)	.43*** (168)	$\alpha = .80$

Note. CE = Coaching effectiveness; WAI = Working Alliance Inventory, WEMWBS = Warwick-Edinburgh Mental Well-Being Scale, MSPSS = Multidimensional Scale of Perceived Social Support; GSE = General Self-Efficacy; BRS = Brief Resilience Scale. All measurements taken at T3, except for GSE and BRS, which were only measured at T1. Sample sizes of the correlations are displayed within parentheses.

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

“there is a special person who is around when I am in need.”

**Mental Well-Being.** Mental well-being was assessed with help of the 14-item Warwick-Edinburgh Mental Well-Being Scale (WEMWBS; Tennant et al., 2007), which covers both psychological and subjective aspects, in other words, satisfaction and happiness, or technically “eudemonic” and “hedonistic” criteria of mental well-being.

**General Self-Efficacy.** The General Self-Efficacy scale was used to assess the coach/coachee self-efficacy (GSE; Schwarzer & Jerusalem, 1995). This scale consists of 10 items on a 4-point Likert scale. Sample items include: “I can always manage to solve difficult problems if I try hard enough,” “If someone opposes me, I can find the means and ways to get what I want,” and “It is easy for me to stick to my aims and accomplish my goals.”

**Resilience.** Resilience was measured with the Brief Resilience Scale (BRS; Smith, Dalen, Wiggins, Tooley, Christopher, Bernard, 2008) which consists of 6 items on a 5-point Likert scale, with sample items “I tend to bounce back quickly after hard times,” and “I usually come through difficult times with little trouble.”

**Coach Behaviors.** For measuring the behaviors of the coach from the viewpoint of coachee and coach, we used the Ashridge Coaching Behaviors Questionnaire (CBQ), which has recently been used in research and validated (De Haan & Nilsson, 2017). The questionnaire has 72 items that are measured in an ipsative way and map 6 different coaching behaviors: Prescribing, Confronting, Informing, Releasing, Exploring, and Supporting. For more information on these coaching behaviors and the CBQ, see the Appendix.

**Personality Characteristics.** For measuring important aspects of personality, we used the Hogan Personality Suite (Hogan & Hogan, 1997). The Hogan Suite was selected not only for its psychometric properties (e.g., validity and reliability), but also for its use of relatively accessible language in the items and feedback information provided. There are three individual personality-related measures. The Hogan Personality Inventory (HPI) is a measure of normal personality based on common “bright-side” dimensions. Responses are forced-choice yes/no. The test–retest reliabilities range from .69 to .87. The Hogan Development Survey (HDS) is an instrument to assess 11 “derailer” or maladaptive behavioral dimensions (Hogan & Hogan, 1997). Respondents indicate to what extent they “agree” or “disagree” with the items. The measure has been cross-

validated with measures of abnormal personality, such as the Minnesota Multiphasic Personality Inventory. The internal reliabilities for the scales are an average alpha of .67 with an average test–retest reliability of .75. The Motives, Values, Preferences Inventory (MVPI) is a measure of 10 motives/preferences that are thought to drive individual behavioral interests, intent, and effective engagement. Respondents indicate to what extent they “agree” or “disagree” with the items. MVPI scores have test–retest reliabilities at an average of .79. For more information on the different scales in the Hogan Personality Suite see the Appendix. In addition to using the Hogan psychometrics, we also asked all coaches and coachees to give us their Myers-Briggs Type Inventory (MBTI; Myers et al., 1998) preferences if they remembered them. 73% of coaches and 43% of coachees did provide us with these scores.

### Data Pre-Analysis Checks

Descriptive statistics, reliability estimates, and intercorrelations for time point T3 are reported in Table 2, except for GSE and BRS, which were only measured at T1. As in other studies employing the same variables, all variables show good internal consistency ( $\alpha > .78$  throughout). The other time points show a similar picture. We tested the comparability of the measure for Coaching Effectiveness (CE) over time; that is, before and after coaching, by establishing longitudinal measurement invariance. Confirmatory Factor Analysis (CFA) is typically viewed as the best approach for this (e.g., Bollen, 1989; Vandenberg & Lance, 2000). A configural invariance model was tested across the three time periods. The fit indices of GFI = .957; NFI = .981; CFI = .977, and SRMR = .0377 can be considered excellent and are all well above accepted threshold levels (e.g., Hu & Bentler, 1999), supporting configural invariance. A metric invariance model was examined next by checking the CFI. Following the suggestions of Cheung and Rensvold (2002) and Chen (2007), the decrease for the CFI value from the configural invariance model to the metric invariance model should not exceed 0.01. The CFI for the metric invariance model is .972, so metric invariance is supported as well.

We also checked for the implications of our multilevel data structure by comparing the spread in effectiveness scores for the 30 coaches having multiple coachees against the overall spread in effectiveness for all coach data. The average SD of the nested data

were 0.55, which is smaller than the average SD of the overall sample (0.83) as expected; however, this difference was only 4% of the mean and therefore insignificant. For this reason, we did not incorporate the nested structure into our analysis.

We checked for any differences between male and female coaches by doing independent *t* tests for CE-scores of coaches and coachees between female–female and female–male pairings. We found no significant results, indicating no gender bias in our sample.

**Data Analysis**

We first established coaching effectiveness (CE) by comparing the development of CE scores between experimental and control groups. Then, on the lookout for active ingredients that have an impact on coaching outcomes and personality, we fitted all data as much as possible to the model in Figure 1. What is unique and different about this model is that we propose a direct positive effect of coaching on personality derailment patterns, as predicted by Nelson and Hogan (2009) and our Hypothesis 6.

Furthermore, as has been suggested before (Baron & Morin, 2009; Boyce et al., 2010; De Haan et al., 2013), we assume that the coaching relationship, coach behaviors, and coachee characteristics, such as well-being, social support, self-efficacy, and personality factors, can all be regressed on outcome, but are to a large extent mediated by the best predictor: the coaching relationship (measured by WAI). This latter aspect of our model has been tested before without a control group (e.g., De Haan et al., 2016). Because of the large number of variables, we applied the Bonferroni correction at this stage. We explored the hypothesized relationship between coaching and personality change by establishing significant changes in (Hogan) personality aspects as a result of coaching; that is, changes that are demonstrable in the experimental group between Time Points 1 and 2, *but not* according to our model in the control group.

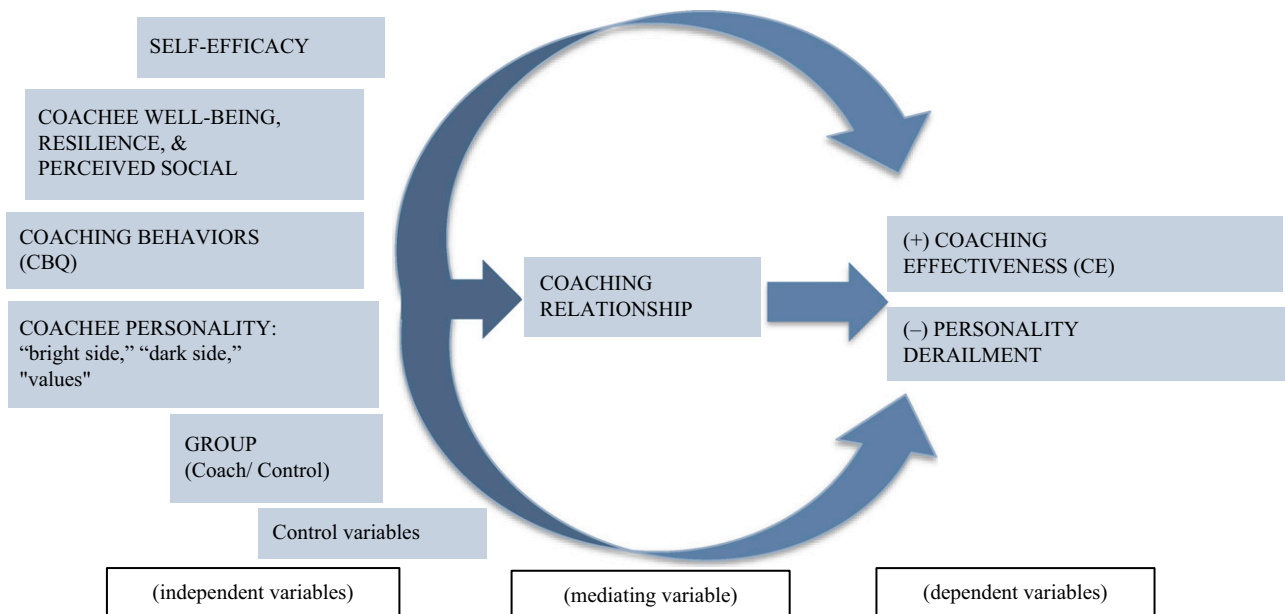
**RESULTS**

*Hypothesis 1: CE Was Demonstrated as Significantly Different From Controls*

**FIGURE 1**

**Our hypothesized model of coaching effectiveness containing only aspects that are common to all coaching conversations and coaching models, with six independent variables to impact on coaching outcome (coaching relationship WAI, coach behaviors, coachee well-being, social support, self-efficacy, and personality factors).**

**There are two hypothesized outcomes: enhanced coaching effectiveness (CE) and reduced personality derailment (i.e., relevant changes in Hogan’s HPI, HDS, MVPI). The model assumes that most of the “common factors” that are significant will be *mediated* through the coaching relationship (measured by WAI).**



Hypothesis 1 was supported from the perspective of the coachees (self-scores) and also from that of their line managers, but the effect from the coaches' perspective did not reach significance. Despite the fact that perspectives on outcome do not correlate between coachee, coach, and line manager CE (Table 2), it is possible to see large effects from both the coachees' and the line-managers' perspectives (see Table 3). We found a substantial difference in outcomes between the two cohorts (Figure 2), reflected in large effect sizes between the two groups: Coachee-estimated CE effect size at TM2 is  $d = 1.08$ , and line-manager-estimated CE effect size at TM2 is  $d = 1.05$ , and still significant at TM3 where the control group has already received half of their coaching:  $d = 0.38$  (see Figure 2). The strength of the effect was larger than anticipated, which may be due to the addition of group coaching and managerial support to pure executive coaching, as part of the overall leadership development program.

*Hypothesis 2: The Strength of the Working Alliance Will Predict CE*

We confirmed Hypothesis 2: There was a strong positive correlation between the coach-reported WAI and CE as measured by the coach, and the same for the coachees' experienced overall WAI and CE as measured by the coach. Moreover, the coachee's WAI also predicted CE as measured by the coachee at a much later time point (see Table 4 for the results for Cohort 1, which were similar for the other

cohort). This is entirely consistent with earlier research in psychotherapy: The client's experience of the working alliance appears the most stable over time, as in psychotherapy (Safran et al., 1993).

Consistent with previous research by Baron and Morin (2009), Boyce et al. (2010), Gessnitzer and Kauffeld (2015), and De Haan et al. (2013, 2016), summarized in a recent meta-analysis (Graßmann, Schölmerich, & Schermuly, 2019), we find some evidence that coachee-rated WAI also predicts coach and line-manager measured CE, but not for all factors and only for the same time point (see Table 4). We would expect that these correlations would become significant with a larger sample size. As in De Haan et al. (2016) some of the correlations between CE and the Bond aspect of the working alliance seem weaker. However, computation of Fisher's  $r$ -to- $z$  transformation showed that none of the differences between task and goal versus bond aspects of the working alliance were significant, possibly because sample size is much smaller here than in De Haan et al. (2016).

*Hypothesis 3: General Self-Efficacy, Resilience, Social Support and Mental Well-Being Will Predict CE*

Hypothesis 3 was supported in that the coachee's aspects of mental preparedness such as self-motivation, perceived social support, resilience, and well-being are also significantly related to CE, but less so than the coachee's experience of the coaching relationship WAI (see Table 5). These

**TABLE 3**  
**Differences Between Coaching Effectiveness (CE) Reported by Experimental and Control Groups**

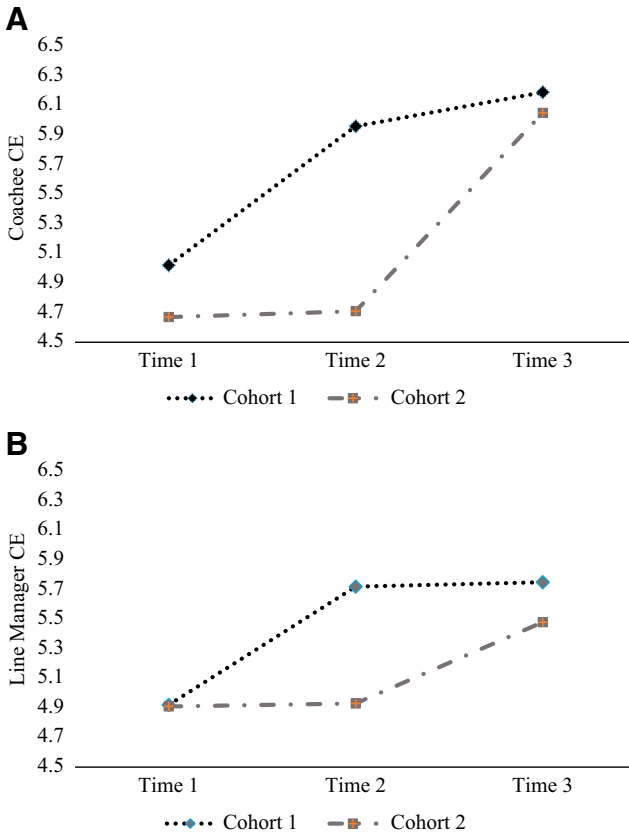
	Interaction						<i>f</i>	<i>p</i>	<i>d</i>
	Cohort 1			Cohort 2					
	T1	T2	T3	T1	T2	T3			
Coachee CE	5.02 (1.11)	5.96 (0.79)	6.19 (0.97)	4.67 (0.96)	4.71 (1.44)	6.05 (0.63)	10.15	<.001	.75
Line Manager CE	4.92 (0.99)	5.72 (0.69)	5.75 (0.60)	4.91 (0.71)	4.93 (0.81)	5.48 (0.80)	3.51	.055	.53

	Simple Effects						Cohort differences (independent t tests)		
	Cohort 1 (estimated mean differences)			Cohort 2 (estimated mean differences)			T1	T2	T3
	T1-T2	T1-T3	T2-T3	T1-T2	T1-T3	T2-T3			
Coachee CE	0.94***	1.17***	0.23	0.04	1.42***	1.38***	0.35	1.25***	0.14
Line Manager CE	0.80*	0.83**	0.03	0.05	0.53*	0.48***	0.18	.57*	.55***

*Note.* Table 3 shows the findings from two Mixed Factorial ANCOVAs, controlling for the gender of the Line Manager, and "preparedness" variables MSPSS, WEMWBS, GSE, and BRS. The columns show means (with SDs in parentheses), *f* values from the interaction effect, *p* values, and effect sizes. The simple effects display mean differences between the time points and the cohorts.  
\* =  $p < .05$ , \*\* =  $p < .01$ , \*\*\* =  $p < .001$ .

**FIGURE 2**  
**Graphical appreciation of differences in Coaching Effectiveness (CE) between control and experimental groups, as perceived by (a) the coaches and (b) their line managers. Cohort 1 is the experimental group and Cohort 2 is the waiting-list control group.**



“preparedness” factors were also significantly related to the working alliance, with similar strengths as in earlier research by De Haan et al. (2013, 2016). We see a marked reduction in predictive value of these dimensions as compared with the WAI scores (compare Tables 3 and 4). There were no consistent significant correlations with coach or line manager scores, only random fluctuations.

If we take the two variables that have been measured at every time point; namely, WEMWBS or MSPSS, then we can demonstrate through mixed factorial ANOVAS that they significantly increase with coaching in the experimental group compared with the control group:

- Mental well-being. There was a significant interaction of WEMWBS development and cohort,  $f(2, 208) = 3.59, p = .029$ , partial eta square = .033.

- Perceived social support. There was a significant interaction of the MSPSS development and cohorts,  $f(2,206) = 4.27, p = .015$ , partial eta square = .040.

As in previous studies (such as Baron & Morin, 2009; Boyce et al., 2010; De Haan et al., 2013, 2016), we looked at mediation according to the model depicted in Figure 1. Because coach–coachee personality differences were weak predictors of CE, and coachee self-efficacy and resilience did not predict WAI, we reduced the model and checked for possible mediation (Baron & Kenny, 1986), regressing perceived social support and mental well-being and aspects of the working alliance (Task, Bond, Goal) on CE separately for the coachee at T2 (see Figure 1).

The mediation model with MSPSS as a predictor variable, WAI as a mediating variable, and CE T2 as outcome variable suggest that MSPSS significantly predicts WAI,  $b = 0.23, t = 2.17, p = .03$ , WAI in turn significantly predicts CE T2,  $b = 0.73, t = 8.55, p < .001$  and MSPSS significantly predicts CE T2 with WAI excluded from the model,  $b = 0.28, t = 2.58, p = .01$ . However, adding the WAI variable into the model takes away the effect of MSPSS on CE T2,  $b = 0.11, t = 1.41, p = .16$ . Furthermore, the significant bootstrapped indirect effect confirms that the effect of coachee Perceived Social Support was *fully mediated* by the working alliance ( $b = 0.17, 95\% \text{ CI } [0.006, 0.353]$ ).

The other mediation model where WEMBS worked as the predictor variable with WAI as a mediating variable and CE T2 as outcome variable suggests that WEMBS significantly predicts WAI,  $b = 0.03, t = 2.56, p = .01$ ; WAI in turn significantly predicts CE T2,  $b = 0.71, t = 8.33, p = .000$  and WEMBS significantly predicts CE T2 with WAI excluded from the model,  $b = 0.04, t = 3.40, p = .001$ . Adding WAI to the model reduces the effect of WEMBS on CE T2; however, the relationship between WEMBS and CE T2 is still significant,  $b = .02, t = 3.40, p = 0.04$ . Even with a significant effect of WEMBS and CE T2, a significant bootstrapped indirect effect indicates the effect of coachee Mental Well-Being was *partially mediated* by the working alliance ( $b = 0.02, 95\% \text{ CI } [0.005, 0.047]$ ).

*Hypothesis 4: None of the Six Classes of Coaching Behaviors Will Predict CE*

Hypothesis 4 was supported. The six coaching behaviors measured at time points 2 and 3 were used in multiple regressions with the six different CE measurements (coachee, coach, and line manager, over two time points; see Table 6). All six CBQ scales were first entered into the regression model, but the

**TABLE 4**  
**Pearson's Correlations Between Coaching Relationship Expressed by Working Alliance (WAI) and Coaching Effectiveness (CE) as rated by Coachees, Coaches and Line Managers, Based on Cohort 1 Only.**

	Coachee WAI					
	1	2	3	4	5	6
<b>Coachee</b>						
WAI – Task T2		.84*** (73)	.90*** (73)	.78*** (58)	.60*** (58)	.70*** (58)
WAI – Bond T2			.81*** (73)	.67*** (58)	.55*** (58)	.58*** (58)
WAI – Goal T2				.72*** (58)	.60*** (58)	.65*** (58)
WAI – Task T3					.86*** (74)	.90*** (74)
WAI – Bond T3						.84*** (74)
<b>Coach</b>						
WAI – Task T2						
WAI – Bond T2						
WAI – Goal T2						
WAI – Task T3						
WAI – Bond T3						
WAI – Goal T3						

	Coach WAI					
	7	8	9	10	11	12
<b>Coachee</b>						
WAI – Task T2		-.01 (43)	.20 (43)	.24 (48)	.15 (48)	-.06 (48)
WAI – Bond T2		.12 (43)	.25 (43)	.12 (48)	.11 (48)	-.09 (48)
WAI – Goal T2		.14 (43)	.23 (43)	.14 (48)	.07 (48)	-.12 (48)
WAI – Task T3		-.05 (40)	.04 (43)	.24 (50)	.14 (50)	-.02 (50)
WAI – Bond T3		.01 (40)	.10 (40)	.15 (50)	.15 (50)	.03 (50)
WAI – Goal T3		-.11 (40)	-.01 (40)	.22 (50)	.18 (50)	-.06 (50)
<b>Coach</b>						
WAI – Task T2			.75*** (61)	.57*** (37)	.31 (37)	.37* (37)
WAI – Bond T2			.52*** (61)	.47*** (37)	.42*** (37)	.39* (37)
WAI – Goal T2				.70*** (37)	.54*** (37)	.51** (37)
WAI – Task T3					.60*** (60)	.63*** (60)
WAI – Bond T3						.46*** (60)
WAI – Goal T3						

	Coachee			Coach			Line Manager		
	CE T2	CE T3		CE T2	CE T3		CE T2	CE T3	
<b>Coachee</b>									
WAI – Task T2	.74*** (73)	.64*** (58)	.30* (46)	.11 (48)	.11 (48)	.11 (48)	.11 (43)	.02 (41)	
WAI – Bond T2	.64*** (73)	.48*** (58)	.34* (46)	.02 (48)	.02 (48)	.02 (48)	.32* (43)	-.07 (41)	
WAI – Goal T2	.71*** (73)	.58*** (58)	.28 (46)	.07 (48)	.07 (48)	.07 (48)	.23 (43)	.13 (41)	
WAI – Task T3	.62*** (60)	.77*** (73)	.13 (43)	.15 (50)	.15 (50)	.15 (50)	.06 (41)	.12 (42)	
WAI – Bond T3	.46*** (60)	.70*** (73)	.11 (43)	.19 (50)	.19 (50)	.19 (50)	.17 (41)	.15 (42)	
WAI – Goal T3	.51*** (60)	.70*** (73)	.12 (43)	.21 (50)	.21 (50)	.21 (50)	.12 (41)	.09 (42)	
<b>Coach</b>									
WAI – Task T2	.37* (44)	.16 (42)	.62*** (60)	.16 (37)	.16 (37)	.16 (37)	.20 (29)	.29 (29)	
WAI – Bond T2	.03 (44)	.23 (42)	.49*** (60)	.37* (37)	.37* (37)	.37* (37)	.09 (29)	.08 (29)	
WAI – Goal T2	.31* (44)	.24 (42)	.60*** (60)	.24 (37)	.24 (37)	.24 (37)	.27 (29)	.16 (29)	
WAI – Task T3	.13 (50)	.04 (50)	.37* (41)	.21 (59)	.21 (59)	.21 (59)	.23 (33)	.41* (33)	
WAI – Bond T3	.05 (50)	.13 (50)	.28 (41)	.25 (59)	.25 (59)	.25 (59)	.15 (33)	.13 (33)	
WAI – Goal T3	.09 (50)	-.14 (50)	.40*** (41)	.20 (59)	.20 (59)	.20 (59)	.21 (33)	.30 (33)	

Note. Sample sizes of the correlations are displayed within the brackets.  
 \* =  $p < .05$ , \*\* =  $p < .01$ , and \*\*\* =  $p < .001$ .

**TABLE 5**  
**Relationships Between Mental Well-Being (WEMWBS), Perceived Social Support (PSE), and General Self-Efficacy (GSE) and Resilience (BRS)**

	Cohort 1 Coachee			Cohort 2 Coachee		
	CE T1	CE T2	CE T3	CE T1	CE T2	CE T3
WEMWBS T1	.13 (89)	.20 (68)	.22 (67)	.28* (72)	.22 (61)	.15 (59)
MSPSS T1	.21* (89)	.16 (68)	.19 (67)	.18 (72)	-.01 (61)	.13 (59)
WEMWBS T2	.14 (67)	.40*** (76)	.24 (61)	.33** (62)	.13 (74)	-.13 (59)
MSPSS T2	.30* (67)	.32** (76)	.37** (61)	-.22 (62)	-.08 (74)	.12 (59)
WEMWBS T3	.07 (17)	.47*** (62)	.46*** (76)	-.01 (58)	.04 (57)	.05 (75)
MSPSS T3	.17 (67)	.51*** (62)	.45*** (75)	.15 (57)	.11 (56)	.07 (74)
GSE	.28** (89)	.14 (68)	.04 (67)	-.02 (72)	-.10 (61)	.06 (59)
BRS	.25* (89)	.09 (68)	.06 (67)	-.06 (72)	-.14 (61)	-.02 (59)

Note. Expectations of Coaching and Confidence in Coaching (at Time Point 1), and Coaching Effectiveness (CE) as Rated by Coachees.  
 \* =  $p < .05$ , \*\* =  $p < .01$  and, \*\*\* =  $p < .001$  (Sample sizes of the correlations are displayed within the parentheses).

Exploring subscale had to be deleted due to unacceptable indicators of multicollinearity. The multiple regression analysis indicated that coachee CBQ explains 21% of the variance in coachee CE at T3, and that coach CBQ explains 27% of the variance in Coachee CE T3; however, these effects were not significant when taking a Bonferroni correction ( $0.05/6 = 0.008$ ) due to the multiple comparisons in consideration. The multiple linear regression equations did not reveal any other significant effects.

However, multiple linear regressions with the CBQ subscales were also calculated to predict WAI and CE for Time 2 and Time 3 (see Table 6), and the remaining model significantly explained 26–33% of the WAI T2 scales and 27% of the WAI Bond T3 scale. Bonferroni corrections ( $0.05/8 = 0.006$ ) resulted in only the T3 correlations dropping below statistical significance. At T2 it looks like coachee-rated Confronting and Supporting positively predict their perceptions of the working alliance, while Prescribing (and possibly Informing) contributes negatively. It will be worth looking at this with larger samples.

The CBQ questionnaire here is based on forced choice, which may explain the difference found with the earlier, cumulative questionnaire in De Haan et al. (2011).

*Hypothesis 5: Only Big-Five Personality Factors Will Predict CE*

Hypothesis 5a was supported. We confirmed De Haan et al. (2013, 2016) who show that MBTI-preferences do not predict coaching outcomes by correlating all four MBTI polarities (E vs. I, N vs. S, T vs. F, and J vs. P) with outcomes, both for coaches and for coachees, and both for coachee- and coach-rated outcomes, at three time points. This yielded

36 low correlations with only one weak significance; namely, coach T vs. F, which correlated with coach-rated outcome at T2, suggesting higher (self-rated) outcomes for coaches with an F-preference ( $p < 0.05$ ). This was however not replicated at T3.

Table 7 has the correlations between Hogan's Personality Inventory and CE as viewed by all three groups of participants. There are consistent significant correlations for Adjustment (T2 from coach and coachee perspectives), Ambition (T1, T2, and T3), Interpersonal sensitivity (T2 from coach and coachee perspectives) and Learning Approach (T2 from coach perspective), confirming Hypothesis 5c. This is consistent with earlier results found for Emotional Stability and Openness (Stewart et al., 2008). The latter also found positive results for Conscientiousness (in terms of Hogan's HPI this would be Prudence). We think the high scores for Conscientiousness, which we are not replicating, could be related to the way Stewart et al. collected data over the Internet, which may have meant a lower response rate of more conscientious people (see also Ward, Meade, Allred, Pappalardo, & Stoughton, 2017). All in all, Hypothesis 5 was confirmed even if relationships between CE and coachee personality, although consistent, were weak.

The two other personality instruments, HDS and MVPI, measuring derailment factors and motives/values/preferences, do not show any effects that are statistically meaningful, which confirms Hypothesis 5b. A few correlations are significant at  $p < 0.05$ , but their frequency is no better than chance level and does not replicate at other time points.

*Hypothesis 6: Executive Coaching Significantly Improves Personality Derailment Patterns*

TABLE 6

Relationships Between Coaching Behaviors Measured by the Coaching Behaviors Questionnaire (CBQ), Rated by Coachees and Coaches, and Coachee-Rated Working Alliance (WAI) and Coaching Effectiveness (CE), as Expressed by Multiple Regression Analysis ( $R^2$ ) With Unstandardized Beta Values Of Coachee- and Coach-Rated CBQ Scales T2, With Coachee-Rated WAI T2 And T3 and CE T2 And T3.

Coachee CBQ	T2 WAI Task	T2 WAI Bond	T2 WAI Goal	T3 WAI Task	T3 WAI Bond	T3 WAI Goal	T2 Coachee CE	T3 Coachee CE
CBQ ( $r^2$ )	.32***	.33***	.26**	.24	.27**	.23*	.12	.21*
Confronting	.04**	.02*	.03	.02	.01	.02	.02	.03*
Informing	-.01	-.01	-.01	-.04	-.07**	-.03	.00	-.05*
Prescribing	-.04*	-.05**	-.04**	-.04	-.04**	-.04*	-.02	-.01
Releasing	.00	-.01	-.01	-.03	-.05*	-.01	.00	-.02
Supporting	.03**	.03	.03*	.01	.02	.02	.03*	.02

Coach CBQ	T2 WAI Task	T2 WAI Bond	T2 WAI Goal	T3 WAI Task	T3 WAI Bond	T3 WAI Goal	T2 Coachee CE	T3 Coachee CE
CBQ ( $r^2$ )	.13	.15	.15	.19	.23	.15	.21	.27*
Confronting	-.08	-.09	-.07	-.02	-.04*	-.01	-.10	-.04*
Exploring	-.09	-.09	-.08	-.03	-.03	-.02	-.11	-.04
Informing	-.09	-.11	-.09	.01	-.03	.01	-.12	-.03
Prescribing	-.11	-.12	-.11	-.02	-.02	-.01	-.15	-.05
Releasing	-.07	-.08	-.06	.03	.11	.03	-.09	.03
Supporting	-.08	-.07	-.06	-.01	-.02	.00	-.10	-.03

Note: T2 regressions had an  $n$  of 73 T3 regressions had an  $n$  of 58. The second part shows multiple regressions of coach CBQ scales T2, Coachee WAI T2 ( $n = 43$ ) and T3 ( $n = 42$ ), Coachee CE T2 and T3 ( $n = 44$ ).

Hypothesis 6 was supported as well. To check for any effects of executive coaching on personality, we asked participants in both cohorts to complete the Hogan personality instruments twice, at Time Points 1 and 2; that is, before the second cohort had received any coaching. First, we tested whether any of the HDS dimensions were significantly different between the two measurements, which was only the

case for one dimension in 56 (7 HPI, 10 MVPI, and 11 HDS, times 2 cohorts); namely, HDS “Diligent,” which became significantly lower in the second measurement for Cohort 2,  $p < 0.05$ .

When the personality differences were correlated with coachee-estimated outcomes as perceived at T2 (i.e., after coaching), we found a few significant effects in the experimental group:

TABLE 7

Relationships Between “Bright Side” Personality Factors (HPI, 1<sup>st</sup> Measurement, At Time Point 1) and Coaching Effectiveness CE as Rated by Coachees, Coaches, and Line Managers Over All Time Points.

	Coachee CE			Coach CE			Line Manager CE		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
Adjustment	.09 (148)	.16 (134)	.14 (127)	—	.43** (53)	.01 (76)	.10 (121)	-.09 (92)	.06 (104)
Ambition	.30*** (148)	.17 (134)	.20* (127)	—	-.03 (53)	.02 (76)	.14 (121)	.06 (92)	.08 (104)
Sociability	.11 (148)	-.03 (134)	-.05 (127)	—	.18 (53)	-.07 (76)	-.06 (121)	-.04 (92)	-.10 (104)
Interpersonal Sensitivity	.16* (148)	.09 (134)	.06 (127)	—	.49*** (53)	.07 (76)	.02 (121)	-.10 (92)	.02 (104)
Prudence	.00 (148)	.06 (134)	.05 (127)	—	.10 (53)	.04 (76)	-.05 (121)	-.10 (92)	-.16 (104)
Inquisitive	-.01 (148)	-.03 (134)	.09 (127)	—	-.08 (53)	-.07 (76)	.09 (121)	-.27** (92)	-.10 (104)
Learning Approach	-.06 (148)	.07 (134)	.17 (127)	—	.35** (53)	.15 (76)	.06 (121)	-.02 (92)	-.11 (104)

Note: Table 7 shows Pearson’s correlations between Hogan Personality Inventory (HPI) factors of coachee and coaching effectiveness (CE) as rated by the coachee, based on both Cohorts 1 and 2.

\* =  $p < .05$ , \*\* =  $p < .01$ , and \*\*\* =  $p < .001$ . (Sample sizes of the correlations are displayed within the brackets)



- (1) HPI Prudence goes up with higher outcome,  $r = 0.28$  ( $n = 54$ ,  $p < 0.05$ )
- (2) HDS Excitable goes down with higher outcome,  $r = -0.29$  ( $n = 54$ ,  $p < 0.05$ )

These are some first indications that executive coaching can have an effect on personality, particularly as these correlations were not observed in the control group. The effect seems consistent because Prudence negatively correlates with Excitable (with  $r = -0.32$  according to the Hogan statistical guide). It seems that effective coaching has a small but significant calming, balancing, and responsibility-enhancing effect on personality. Because we use a measure that stretches over both time points, we know that the significant correlations found indicate causality; that is, executive coaching progressively makes a difference to personality measures. However, we need to remain cautious about these results, as applying the Bonferroni correction would make them both insignificant.

## DISCUSSION

The analysis confirms Hypotheses 1 to 6 on the impact of “common factors” on CE. We have found strong indications that the coaching relationship WAI as rated by both the coach and coachee correlates with coach- and coachee-rated CE to a considerable degree. This confirms work that was done by Baron and Morin (2009), Boyce et al. (2010), De Haan et al. (2013, 2016) and others as summarized in Graßmann et al. (2019), while Gessnitzer and Kauffeld (2015) have also demonstrated a significant correlation between observer ratings of the working alliance in the first session and coachee-rated outcome after the final session. Similar to Gessnitzer and Kauffeld, we have found strong indications that WAI as rated by the coachee predicts CE estimates at later times (see Table 4). We have also found indications that coachee self-efficacy, psychological well-being, resilience, and social-support all correlate with coach- and coachee-rated CE, and that psychological well-being and social support themselves grow as a result of coaching. Finally, we have found very little evidence for a differential impact of specific coaching behaviors on coach- or coachee-rated CE. This means our results are aligned with those summarized by Graßmann et al. (2019) regarding the coaching relationship; those by Stewart et al. (2008) in the area of self-efficacy; those by Grant et al. (e.g., 2009, 2010) related to well-being even if theirs is a different; namely, more work-related, well-being scale; and those by De Haan et al. (2011) regarding coaching behaviors.

## Confirmation That Executive Coaching Can Be an Effective Intervention

Figure 2 gives us strong evidence that executive coaching can be an effective intervention; not only in the eyes of the coachees, but also in the eyes of their line managers (see Figure 2b), confirming meta-analysis studies such as Jones et al. (2015). By using a randomized control group, we know that the effectiveness can be attributed to the intervention itself. Moreover, thanks to measuring effectiveness through the eyes of coaches, coachees, and line managers, we know that the finding is robust against same-source bias (admittedly, to an extent these three parties will influence each other; however, they have given their estimates entirely independently). The fact that we have found such large effect sizes ( $d$  larger than 1) may have something to do with the waiting-list control group design (e.g., MacKie, 2014, found very similar effect sizes with such a design). Although the study was undertaken within the healthcare industry (as in Grant et al., 2009), we believe that these findings are globally generalizable over many industries, because the coachees were globally mobile, senior- and mostly general managers, not technical healthcare experts.

## Active Ingredients Such as Coaching Relationship, Coachee “Preparedness,” and Personality Are Confirmed

We suggest that industrial buyers can feel reassured that coaching is generally a very effective intervention as reviewed by coachees and their line managers, in such a way that it may warrant investment for development, but also that this effectiveness depends on a strong match between coach and coachee; that is, a strong coaching relationship. Because the working alliance as rated by the coachee also predicts estimates at later times (see Table 4), we maintain that matching is best done by the coachee him- herself, on the basis of a chemistry meeting or trial session and not by HR or other intermediaries (more evidence that demographics or external matching make no difference to CE can be found in Boyce et al., 2010; Page & De Haan, 2014; and Bozer et al., 2015).

We have also found a significant (even if smaller) prediction of outcomes by coachee-related factors such as resilience, self-efficacy, perceived social support, and mental well-being, all factors indicating a good “preparedness” for the impact of coaching (this extends earlier results by Stewart et al., 2008; and De Haan et al., 2013, 2016, which focused on self-efficacy only). Finally, we confirm the relationship

between coachee personality and outcome for some well-researched personality factors (namely Adjustment, Ambition, and Interpersonal sensitivity, see Table 7, confirming results by Stewart et al., 2008). These findings are relatively small and not consistent over all times and sources, so it is too early to try and interpret them.

### First Indications That Coaching May Play a Role in Preventing Leadership Derailment

Even if only as a first indication, we can report that executive-coaching effectiveness as seen by the coachee was significantly correlated with changes in two personality measures; namely, Prudence and Excitable. Prudence rises significantly while Excitable lowers in the experimental group. This may be a single, coherent impact of coaching because the two measures are inversely related (with a correlation of about  $-0.32$ , see Hogan & Hogan, 1997). If we look at the meaning of these measures (see Appendix), we find that the leaders being coached are becoming *more* self-disciplined, responsible, or conscientious while becoming *less* moody, annoyed, hard to please, or emotionally volatile. Executive coaching seems to have a demonstrable calming or containing effect, which is exactly what it sets out to do in being a conversational, reflective practice. This result confirms a large-scale meta-analysis study into the effect of psychotherapy on personality traits, which has reported midsize effects ( $d = 0.37$ ) that persist over substantial periods after treatment of similar personality traits, such as emotional stability and extraversion (Roberts, Luo, Briley, Chow, Su, & Hill, 2017).

At this point, we do not know if coaching does have the capacity to change the bright or dark sides of personality, or if coaching only affects mood states in such a way that the coachee only “feels” better or calmer. We also do not know which aspects of coaching relationships help most to produce the measured changes. All we know is that measurable adjustments on rigorous psychometrics do occur in our sample. The case could be that coachees are generally calmer or uplifted by the coaching sessions; however, in that case, we would have expected more effects on other personality dimensions as well. Moreover, Roberts et al. (2017) make a plausible case that in their very large sample ( $N = 20,024$ ), personality traits have been changed substantially and robustly, a change that could not be explained by mood changes.

Another possibility is that coachees may have received new insight that sensitized them to their habitual responses, so that they became more self-

aware, adaptive, or self-regulated in general ways, without fundamentally changing their deeper “personalities.” We argue this would still change their reputation in the workplace in a positive manner, so at least their personalities as perceived by others would be affected positively. Moreover, the tools we used, the HPI and HDS, are well known for measuring lasting reputations in the workplace.

Another area for speculation and in the future for possible research is asking what exactly contributes to the realized changes in these “personality derailers?” We conjecture that to make an impact on these deeper levels of personality in such a short time frame, a combination of substantial safety, trust, and intimacy with direct, frank, and robust challenge would have been required. This may help us to conceptualize coaching for leaders as a deeply *supporting* (or, more technically, “containing”) and robustly *challenging* intervention. Interestingly, we have also found *Supporting* and *Confronting* to be the only coach behaviors scored by the coachees as significantly and positively correlated with their WAI scores (see Table 6 for T2). If support and challenge are indeed important, that finding would argue against purely *facilitative* coaching approaches such as those promoted by prominent coach associations (e.g., the International Coach Federation) and coaching handbooks (e.g., Whitmore, 1992).

### SUMMARY

In summary, we believe this research can help potential buyers of coaching (HR, line managers, and Learning & Development professionals) in four different ways: first, executive coaching appears to be a significantly effective intervention in the coachees’ view, as well as independently in their line managers’; second, some of this effectiveness has an impact on deeper layers of the leader’s personality; third, effectiveness depends to a large extent on the strength of the working alliance between coach and coachee; and fourth, coaching effectiveness can be further improved by coachees being mentally prepared, resilient, and motivated, while coaching in turn improves their well-being and sense of support in a virtuous circle.

### LIMITATIONS

Some limitations of our study have to be acknowledged. First, participants experienced both individual and group coaching, and some had a few conversations with senior-level managers, which pose

potentially confounding effects on the outcomes. Hence, some of the CE gains may have arisen from additional management support, rather than the impact of the coaching itself. Furthermore, even though we asked explicitly about the effects of *individual* coaching on outcomes, the reported effects may just as well have been reached by group coaching as by individual coaching. At least we can guarantee that no training or teaching took part within this organizational intervention, so we are convinced that effects found should be ascribed to coaching (and perhaps management conversations prompted by the program).

Second, same-source bias is expected in some of the reported effects (when coachee measures are compared). Even when different sources are employed (e.g., coaches and line managers), one has to bear in mind that these other sources were both contributing to the intervention, for example, at the triangular meetings beginning and ending the coaching, so they were not fully independent.

Third, the sample sizes of some measurements were lower than would be hoped, mainly because of the attrition rate of responses from both the experimental and control groups, and from coaches and line managers in particular. As Ward et al. (2017) suggest, attrition (and careless responses, which may be related to attrition) are potential sources of bias in online surveys. Their two studies found attrition and careless responses relating to the Big-5 personality traits and to bias. Conscientiousness, Extraversion, Emotional Stability, and Agreeableness were significantly related to survey attrition and carelessness resulting in significant differences between respondents who carefully completed the survey and those lost by way of attrition or careless response screening (Ward et al., 2017). However, they also find that carelessness is more of an overall trait and does not vary at different time points. Moreover, the significant effects of attrition only occur in Extraversion and Emotional Stability, and for Emotional Stability the effect is actually *opposite* the one we found for Prudence and Excitable. Ward et al. (2017) found that those who dropped out in their study are lower in neuroticism, so one would imagine a less prudent and more excitable population to remain in the study. Our effects reported above are exactly in the opposite direction.

Fourth, there are small risks of distortions through researching exclusively female leaders and through nested data: Some coaches had more than one coachee, and some line managers had more than one direct-report on the program. Also, there may be

nesting effects through geographies or divisions of the company. After checking for multilevel effects, we think these risks are small. Moreover, there is no evidence in the literature for any sizable differences between men and women as coachees (the only small, but significant differences that we know of are in Bozer et al. 2015, and De Haan et al., 2016, and we did check for gender effects in the coaching relationships, but could not demonstrate any).

Fifth, there are limitations as to the generalizability of the study sample to managers in other positions, organizations, industries, and geographical regions. Although we believe the kind of roles and responsibilities of the women leaders in the sample are common to managerial roles in other large organizations, we acknowledge that other settings may contain numerous unique factors that influence the effectiveness of coaching interventions. Furthermore, while our focus on women only enabled us to control for the gender variable, it may limit the generalizability of the results to all managers, including male managers.

## IMPLICATIONS AND FUTURE DIRECTIONS

### Implications for Theory

More coaching theory is needed to conceptualize how this highly tailored, empathetic, supportive, and yet challenging, thoroughly confidential intervention might have an impact on those deeper levels of the personality. What is it about coaching that makes it such a promising and effective leadership-development tool in the presence of pressures such as high anxiety, high stress, or powerful projections? Senior managers are most prone to those pressures, as Kaiser, LeBreton, and Hogan (2015) show by quoting studies demonstrating that personality-derailment risks (as measured by the Hogan HDS) are more prevalent in higher management cadres. Kaiser and Kaplan (2006) suggest that coaches must be prepared to confront the subtle fears of failure, inadequacy, and rejection that can cloud the executive's judgment and impair interactions with subordinates and peers. We need more theory to model and then test this kind of "tough love" that combines the offer of a safe confidential space with considerable challenge. However, given the methodological issues we encountered in exploring personality factors, we recommend that future research continues exploring this issue using larger samples. We need more theory about the impact of effective coaching on, for example, performance and personality change of the leader and how this affects the wider organization,

following initial qualitative studies such as, for example, by Swart and Harcup (2013). Levenson (2009) argues that even if coaching may be demonstrated to work on personality derailment (such as is the case with our sample), that does not necessarily demonstrate a positive impact on business results or organizational context. So, what is needed is more research, but also crisp, verifiable theory that brings management stress and anxiety, executive coaching outcomes, well-being, resilience, mindfulness, personality derailers, and performance measures in the workplace together in a single, testable model similar to our Figure 1. Such a model could be based on the large, existing research body regarding the impact of mindfulness through well-being on performance (see for a recent example and summary relevant to management education, Lomas, Medina, Ivtzan, Rupprecht, & Eiroa-Orosa, 2017).

### Implications for Practice

These findings extend past work to provide research that may help guide the coaching industry and the development of evidence-based approaches to coaching, as well as inform the choices that are made in the recruitment, development, deployment, and matching of executive coaches (as e.g., summarized in Wycherley & Cox, 2008). This study clearly demonstrates effectiveness of coaching on objective measures with the help of a randomized controlled trial design. Given this effectiveness, the added costs and time associated with executive coaching may be viewed by organizations as acceptable. However, we agree with Nieminen et al. (2013) that the potentially longer time horizons associated with such interventions may not be appropriate in organizations with more immediate leadership needs and a wider target population. On a practical level, coaches can learn to look into their coachee's "preparedness" for coaching: their motivation, support, and mental well-being generally, and in the initial contracting conversations, advise coachees to look after or be aware of those aspects as well. Second, coaches should keep inquiring into the quality of the working alliance, particularly from the perspective of their coachees, as it is such an important predictor of effectiveness. Third, coaches would do well to develop their support and challenge, and perhaps not be afraid of these more personal, yet risky interventions, because coachees do relate these to a stronger alliance, and they may also be related to the helpful effects we have found on the leader's personality. After all, executive coaching (together with psychotherapy

and counseling, see Roberts et al., 2017) may be the *only* intervention at our disposal within the broader field of leadership education and organizational development, that is customized ("tailored"), supportive, and challenging enough to develop the leader's personality and reputation in the workplace to make a real, meaningful difference for many colleagues and clients, and even entire teams and organizations, at that relatively deep, personal level.

### Future Research

There is still a lively debate in the coaching profession around the question of whether coaching is effective at all (see e.g., Briner, 2012). So first, we recommend more and larger scale RCT studies to (dis-) confirm and deepen the findings of this study. Although one of the unique contributions here was measurement at three time points, we recommend that in future studies more time points, including follow-up measurements, are considered so that the long-term predictive value of the active ingredients can be assessed. This would also help in establishing whether (a) independent variables such as "preparedness" and "working alliance" only drive *baseline levels* of effectiveness (the so-called intercept); that is, general coachability, or do they also have an impact on the *slope of the learning curve* through coaching, the actual added value of coaching per session, and whether (b) there are key turning points in sessions that lead to overall effectiveness and where they might lie in the coaching journey. Second, we need more studies exploring the nature of the demonstrated changes through coaching. This could help to find out more about the directionality of the effect (e.g., with self-efficacy, well-being, and perceived social support we now know that the effect is probably bidirectional). Third, we recommend better (personality-based as well as performance-related) outcome variables to make this demonstration more convincing. Fourth, we recommend that future studies start unpacking the differential effects of group and individual coaching that we had to conflate somewhat here. This would be interesting to do for organizational context and culture as well, to check if there is a moderating factor in terms of the industry, culture, or nationalities participating in the coaching setting.

### REFERENCES

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

- Anderson, S. L., & Betz, N. E. 2001. Sources of social self-efficacy expectations: Their measurement and relation to career development. *Journal of Vocational Behavior*, 58: 98-117.
- Armstrong, S. J. 2011. From the Editors: Continuing our quest for meaningful impact on management practice. *Academy of Management Learning & Education*, 19(2): 181-187.
- Baron, L., & Morin, L. 2009. The coach-coachee relationship in executive coaching: A field study. *Human Resource Development Quarterly*, 20: 85-106.
- Baron, R. M., & Kenny, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173-1182.
- Blackman, A., Morsardo, G., & Gray, D. E. 2016. Challenges for the theory and practice of business coaching: A systematic review of empirical evidence. *Human Resource Development Review*, 15(4): 459-486.
- Bollen, K. A. 1989. *Structural equations with latent variables*. New York: Wiley.
- Boyce, L. A., Jackson, R. J., & Neal, L. J. 2010. Building successful leadership coaching relationships: Examining impact of matching criteria in a leadership coaching program. *Journal of Management Development*, 29: 914-931.
- Bozer, G., Baek-Kyoo, J., & Santora, J. C. 2015. Executive coaching: Does coach-coachee matching based on similarity really matter? *Consulting Psychology Journal*, 67(3): 218-233.
- Bozer, G., Sarros, J. C., & Santora, J. C. 2013. The role of coachee characteristics in executive coaching for effective sustainability. *Journal of Management Development*, 32(3): 277-294.
- Briner, R. B. 2012. Does coaching work and does anyone really care? *Occupational Psychology Matters*, 16: 4-11.
- Bryson, A., Forth, J., & Stokes, L. 2014. Does worker wellbeing affect workplace performance? Department for Business Innovation & Skills. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/366637/bis-14-1120-does-worker-wellbeing-affect-workplace-performance-final.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/366637/bis-14-1120-does-worker-wellbeing-affect-workplace-performance-final.pdf)
- Chen, F. F. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3): 464-504.
- Cheung, G. W., & Rensvold, R. B. 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2): 233-255.
- Dahling, J. J., Taylor, S. T., Chau, S. L., & Dwight, S. A. 2016. Does coaching matter? A multilevel model linking managerial coaching skill and frequency to sales goal attainment. *Personnel Psychology*, 69: 863-894.
- De Haan, E. 2008. *Relational coaching – Journeys towards mastering one-to-one learning*. Chichester: Wiley.
- De Haan, E., & Duckworth, A. 2013. Signaling a new trend in coaching outcome research. *International Coaching Psychology Review*, 8(1): 6-20.
- De Haan, E., & Nilsson, V. 2017. Evaluating coaching behavior in managers, consultants and coaches: A model, questionnaire, and initial findings. *Consulting Psychology Journal*, 69(4): 315-333.
- De Haan, E., Bertie, C., Day, A., & Sills, C. 2010. Critical moments of clients of coaching: Towards a 'client model' of executive coaching. *Academy of Management Learning & Education*, 5(2): 109-128.
- 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., 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*, 65(1): 40-57.
- De Haan, E., Grant, A., Burger, Y., & Eriksson, P.-O. 2016. A large-scale study of executive coaching outcome: The relative contributions of working relationship, personality match, and self-efficacy. *Consulting Psychology Journal: Practice and Research*, 68(3): 189-207.
- Digman, J. M. 1990. Personality structure: Emergence of the Five Factor Model. *Annual Review of Psychology*, 41: 417-440.
- Duijts, S. F. A., Kant, I., Van den Brandt, P. A., & Swaen, G. M. H. 2008. Effectiveness of a preventive coaching intervention for employees at risk for sickness absence due to psychosocial health complaints: Results of a randomized controlled trial. *Journal of Occupational and Environmental Medicine*, 50(7): 765-776.
- Ely, K., Boyce, L. A., Nelson, J. K., Zaccaro, S. J., Hernez-Broome, G., & Whyman, W. 2010. Evaluating leadership coaching: A review and integrated framework. *The Leadership Quarterly*, 21(4): 585-599.
- Evers, W. J. G., Brouwers, A., & Tomic, W. 2006. A quasi-experimental study on management coaching effectiveness. *Consulting Psychology Journal*, 58: 174-182.
- Gessnitzer, S., & Kauffeld, S. 2015. The working alliance in coaching: Why behavior is the key to success. *The Journal of Applied Behavioral Science*, 51(2): 177-197.
- Grant, A. M. 2003. Towards a psychology of coaching: The impact of coaching on metacognition, mental health and goal attainment. *Dissertation Abstracts International Section B*, 63: 6094 and downloadable at <http://files.eric.ed.gov/fulltext/ED478147.pdf>.
- 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 well-being: A randomized 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*, 62(3): 151–168.
- Gray, D. E., & Goregaokar, H. 2010. Choosing an executive coach: The influence of gender on the coach-coachee matching process. *Management Learning*, 41(5): 525–544.
- Graßmann, C., Schölmerich, F., & Schermuly, C. C. In Press. The relationship between working alliance and client outcomes in coaching: A meta-analysis. *Human Relations*.
- Green, L. S., Grant, A. M., & Rynsaardt, J. 2007. Evidence-based life coaching for senior high school students: Building hardiness and hope. *International Coaching Psychology Review*, 2(1): 24–32.
- Green, L. S., Oades, L. G., & Grant, A. M. 2006. Cognitive-behavioral, solution-focused life coaching: Enhancing goal striving, well-being, and hope. *The Journal of Positive Psychology*, 1(3): 142–149.
- 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.
- Hogan, R., & Hogan, J. 1997. *Hogan development survey manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., & Kaiser, R. B. 2005. What we know about leadership. *Review of General Psychology*, 9(2): 169–180.
- Hooijberg, R., & Lane, N. 2009. Using multisource feedback coaching effectively in executive education. *Academy of Management Learning & Education*, 8(4): 483–493.
- Horvath, A. O., & Greenberg, L. S. 1986. The development of the Working Alliance Inventory: A research handbook. In L. S. Greenberg & W. Pinsoff (Eds.), *Psychotherapeutic processes: A research handbook*. New York: Guilford Press.
- Horvath, A. O., & Marx, R. W. 1990. The development and decay of the working alliance during time-limited counselling. *Canadian Journal of Counselling*, 24: 240–259.
- Horvath, A. O., & Symonds, B. D. 1991. Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 61: 561–573.
- Hu, L., & Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(4): 1–55.
- ICF 2016. *Global Coaching Study - International Coach Federation ICF*. [Accessed 7 October 2017 at <https://coachfederation.org/2016study>]
- Jones, R., Woods, S., & Guillaume, Y. 2015. The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, 89(2): 249–277.
- Kaiser, R. B., & Kaplan, R. E. 2006. The deeper work of executive development: Outgrowing sensitivities. *Academy of Management Learning & Education*, 5: 463–483.
- Kaiser, R. B., LeBreton, J. M., & Hogan, J. 2015. The dark side of personality and extreme leader behavior. *Applied Psychology*, 64(1): 55–92.
- Kenworthy, A. L., Passarelli, A., & Van Oosten, E. 2014. Introduction: Coaching and positive emotions: Exploring a collection of resources in the health and wellness domain. *Academy of Management Learning & Education*, 13(2): 290–292.
- Kets de Vries, M. F. R. 2006. *The leader on the couch: A clinical approach to changing people and organizations*. Chichester, UK: Wiley.
- 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.
- Levenson, A. 2009. Measuring and maximizing the business impact of executive coaching. *Consulting Psychology Journal*, 61(2): 103–121.
- Liu, X., & Batt, R. 2010. How supervisors influence performance: A multilevel study of coaching and group management in technology-mediated services. *Personnel Psychology*, 63: 265–298.
- Lomas, T., Medina, J. C., Ivtzan, I., Rupperecht, S., & Eiroa-Orosa, F. J. 2017. The impact of mindfulness on the wellbeing and performance of educators: A systematic review of the empirical literature. *Teaching and Teacher Education*, 61: 132–141.
- Luszczynska, A., Scholz, U., & Schwarzer, R. 2005. The general self-efficacy scale: Multicultural validation studies. *The Journal of Psychology*, 139(5): 439–457.
- MacKie, D. 2014. The effectiveness of strength-based executive coaching in enhancing full range leadership development: A controlled study. *Consulting Psychology Journal*, 66(2): 118–137.
- Martin, D. J., Garske, J. P., & Davis, M. K. 2000. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68: 438–450.
- McDermott, M., Levenson, A., & Newton, S. 2007. What coaching can and cannot do for your organization. *Human Resource Planning*, 30(2): 30–37.

- 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.
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. 1998. *MBTI manual*. Palo Alto, CA: Consulting Psychologists Press.
- Nelson, E., & Hogan, R. 2009. Coaching on the dark side. *International Coaching Psychology Review*, 4: 7–19.
- Nieminen, L. G., Smerek, R., Kotrba, L., & Denison, D. 2013. What does an executive coaching intervention add beyond facilitated multisource feedback? Effects on leader self-ratings and perceived effectiveness. *Human Resource Development Quarterly*, 24(2): 145–176.
- O'Neil, D. A., Hopkins, M. M., & Bilimoria, D. 2015. A framework for developing women leaders: Applications to executive coaching. *The Journal of Applied Behavioral Science*, 51(2): 253–276.
- Page, N., & De Haan, E. 2014. Does coaching work? . . . and if so, how? *The Psychologist*, 27(8): 582–586.
- Passmore, J., & Rehman, H. 2012. Coaching as a learning methodology: A mixed methods study in driver development. A randomized controlled trial and thematic analysis. *International Coaching Psychology Review*, 7(2): 166–184.
- Passmore & Theeboom T. 2015. Coaching psychology: A journey of development in research. In L. E. Van Zyl, M. W. Stander, & A. Odendal (Eds.), *Coaching psychology: Meta-theoretical perspectives and applications in multi-cultural contexts*. New York, NY: Springer.
- Passmore, J., & Velez, M. J. 2012. Coaching fleet drivers: A randomized controlled trial (RCT) of short coaching interventions to improve driver safety in fleet drivers. *The Coaching Psychologist*, 8(1): 20–26.
- Ridler Report. 2016. The Ridler Report 2016. Available <http://www.ridlerandco.com/ridler-report/>. (Last Access 12/11/2019).
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. 2017. A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143(2): 117–141.
- Roehrle, B., & Strouse, J. 2008. Influence of social support on success of therapeutic interventions: A meta-analytic review. *Psychotherapy (Chicago, Ill.)*, 45(4): 464–476.
- Safran, J. D., Crocker, P., McMains, S., & Murray, P. 1990. Therapeutic alliance rupture as a therapy event for empirical investigation. *Psychotherapy (Chicago, Ill.)*, 27: 154–165.
- Safran, J. D., Muran, C. J., & Wallner Samstag, L. 1993. Resolving therapeutic ruptures: A task analytic investigation. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice*. New York: Wiley.
- Sherer, M., Maddux, J. E., Mercandante, B., Dunn, P., Jacobs, B., & Rogers, R. W. 1982. The self-efficacy scale: Construction and validation. *Psychological Reports*, 51: 663–671.
- Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., & Zhang, J. X. 1997. The assessment of optimistic self-beliefs: Comparison of the German, Spanish, and Chinese versions of the general self-efficacy scale. *Applied Psychology*, 46(1): 69–88.
- Schwarzer, R., & Jerusalem, M. 1995. Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston, (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs*: 35–37. Windsor, UK: NFER-NELSON.
- Scoular, A., & Linley, P. A. 2006. Coaching, goal-setting and personality type. What matters? *The Coaching Psychologist*, 2: 9–11.
- Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. 2008. The brief resilience scale: assessing the ability to bounce back. *International Journal of Behavioral Medicine*, 15(3): 194–200.
- 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: 23–44.
- Spence, G. B., & Grant, A. M. 2007. Professional and peer life coaching and the enhancement of goal striving and well-being: An exploratory study. *Journal of Positive Psychology*, 2(3): 185–194.
- Stewart, L. J., Palmer, S., Wilkin, H., & Kerrin, M. 2008. The influence of character: Does personality impact coaching success? *International Journal of Evidence Based Coaching and Mentoring*, 6(1): 32–42.
- Swart, J., & Harcup, J. 2013. 'If I learn do we learn?': The link between executive coaching and organizational learning. *Management Learning*, 44(4): 337–354.
- Tennant, R., et al. 2007. The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, 5: 63.
- Theeboom, T., Beersma, B., & van Vianen, A. E. M. 2014. Does coaching work? A metaanalysis on the effects of coaching on individual level outcomes in an organizational context. *The Journal of Positive Psychology*, 9(1): 1–18.
- Vandenberg, R. J., & Lance, C. E. 2000. A review and synthesis of the measurement invariance literature:

Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1): 4–70.

- Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. 2017. Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, 76: 417–430.
- Whitmore, J. 1992. *Coaching for performance: GROWing people, performance and purpose*. London: Nicholas Brealey Publishing.
- Wycherley, I. M., & Cox, E. 2008. Factors in the selection and matching of executive coaches in organisations. *Coaching (Abingdon, UK)*, 1: 39–53.
- Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. 1988. The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, (52): 30–41.



**Erik de Haan** is the director of the Ashridge Centre for Coaching at Hult International Business School and Professor of Organisation Development and Coaching at the VU University Amsterdam. He has published nearly 200 professional and research articles and 12 books, covering his main fields of expertise as a leadership and organisational consultant, facilitator, supervisor, and coach. He has an MSc in Theoretical Physics, an MA in Psychodynamic Psychotherapy, and a PhD in Psychophysics; and specializes in team coaching and one-to-one coaching for executives.

**David Gray** (now deceased) was a professor of Leadership and Organisational Behaviour at the University of Greenwich. His research interests, and publication record, included research methods, management learning, professional identity, action learning, reflective learning, management learning in Small and Medium-sized Enterprises (SMEs), and the factors that contribute to SME success. He published books and articles on research methods, organizational learning, coaching and mentoring.

**Sally Bonneywell** is an independent executive coach, consultant and researcher. She obtained her doctorate in Coaching and Mentoring from The International Centre of Coaching at Oxford Brookes University, Oxford. She is based in London and works with client organisations to enable people, teams, and organizations to flourish. She is a Lead Coach at IMD in Lausanne, and also coaches executives through Said Business School, Oxford University. Her current research interests include the use of different coaching modalities, including writing coaching, changes in the coaching industry, and achieving sustainable change through narrative coaching.



## APPENDIX

A more detailed description of all dimensions in the Coaching Behaviors Questionnaire and the Hogan Personality Suite:

- (1) The Ashridge Coaching Behaviors Questionnaire is based on a model of “six categories of interventions” proposed by John Heron and looks at coaching skills and behaviors in a broad sense:
  - a. *Prescribing*: giving directions, advice, and recommendations to the learner/client.
  - b. *Informing*: giving information and knowledge to the learner/client.
  - c. *Confronting*: challenging the learner/client’s assumptions; stimulating their awareness of their own behavior, attitudes, or beliefs.
  - d. *Releasing*: helping the learner/client to release tension, and to discharge or come to terms with emotions that are blocking progress.
  - e. *Exploring*: helping the learner/client to self-discovery, to self-directed learning, and to owning and solving his or her own problems, without becoming involved in the learning or change oneself as a coach.
  - f. *Supporting*: building the learner/client’s self-esteem, self-confidence, and self-respect.
- (2) The Hogan Personality Inventory (HPI) describes normal, or bright-side personality qualities that describe how we relate to others when we are at our best. This questionnaire uses forced-choice 206 items to map 7 scales:
  - a. Adjustment: confidence, self-esteem, and composure under pressure
  - b. Ambition: initiative, competitiveness, and desire for leadership roles
  - c. Sociability: gregariousness, and need for social interaction
  - d. Interpersonal Sensitivity: tact, perceptiveness, ability to maintain relationships
  - e. Prudence: self-discipline, responsibility and conscientiousness
  - f. Inquisitive: imagination, curiosity, and creative potential
  - g. Learning Approach: achievement-oriented, up-to-date on business matters
- (3) The Hogan Development Survey (HDS) focuses on a person’s tendencies when under stress, to answer the question: “What tendencies could derail this individual’s career or performance?” This questionnaire uses 168 items to map 11 derailment tendencies:
  - a. Excitable: moody, easily annoyed, hard to please, emotionally volatile



- b. Skeptical: distrustful, cynical, sensitive to criticism, focused on the negative
  - c. Cautious: unassertive, resistant to change, slow to make decisions
  - d. Reserved: aloof, indifferent to the feeling of others, uncommunicative
  - e. Leisurely: overtly cooperative, but privately irritable, stubborn, uncooperative
  - f. Bold: overly self-confident, arrogant, inflated feelings of self-worth
  - g. Mischievous: charming, risk-taking, limit-testing and excitement-seeking
  - h. Colorful: dramatic, attention-seeking, interruptive, poor listening skills
  - i. Imaginative: creative, but thinking and acting in unusual or eccentric ways
  - j. Diligent: meticulous, precise, hard to please, tends to micromanage
  - k. Dutiful: eager to please and reluctant to act independently
- (4) The Motives, Values, Preferences Inventory (MVPI), which describes personality from the inside: the core

- goals, values, drivers, and interests that determine what we desire and strive to attain. This questionnaire uses 200 items to map 10 personal motivators:
- a. Recognition: responsive to attention, approval, and praise
  - b. Power: desiring success, accomplishment, status, and control
  - c. Hedonism: orientated for fun, pleasure, and enjoyment
  - d. Altruistic: wanting to help others and contribute to society
  - e. Affiliation: enjoying and seeking out social interaction
  - f. Tradition: dedicated to strong personal beliefs
  - g. Security: needing predictability, structure, and order
  - h. Commerce: interested in money, profits, investment, and business opportunities
  - i. Aesthetics: needing self-expression, concerned over look, feel, and design of work products
  - j. Science: wanting knowledge, research, technology, and data