IS IT TIME TO THINK DIFFERENTLY ABOUT ACTIVE INGREDIENTS IN COACHING?

New findings on the effectiveness of the coaching relationship:

time to think differently about active ingredients?

–

Erik de Haan*, Joanna Molyn & Viktor O. Nilsson

Ashridge Centre for Coaching @ Hult International Business School, UK

Greenwich Business School

*Corresponding author
Erik de Haan
Director of Ashridge Centre for Coaching, Hult International Business School,
Berkhamsted, Hertfordshire, UK. HP4 1NS

Email: erik.dehaan@ashridge.org.uk

Joanna Molyn
Business Faculty, University of Greenwich, London, UK. SE10 9LS

Email: J.Molyn@greenwich.ac.uk

Viktor Olof Nilsson
Hult International Business School,
Berkhamsted, Hertfordshire, UK. HP4 1NS

Email: Viktor.Nilsson@ashridge.hult.edu

Abstract

This article critically reviews two recent, large-scale, randomized controlled trials in executive coaching, in order to drive further exploration into the topic of the *coaching relationship* as a predictor of coaching outcome. One of the trials was designed at senior levels in an industrial setting and the other was an experiment with coaching in a business-school context. Each trial demonstrated considerable and significant coaching effectiveness with the coaching relationship ('working alliance') as an important ingredient of effectiveness. The more recent randomized-controlled-trial sample, which was longitudinal, seems to show that we may have to radically change our

understanding of the impact of the coaching relationship on coaching effectiveness. Contrary to previous consensus, it seems the working alliance between client and coach is not strongly related to coaching effectiveness. The strength of the working alliance only correlates with a higher effectiveness score from the beginning of the coaching relationship, but it does not significantly correlate with increasing outcomes through further coaching conversations. Some possible explanations for this unexpected and seemingly contradictory finding in the area of 'working alliance' are put forward and critically reviewed.

IS IT TIME TO THINK DIFFERENTLY ABOUT ACTIVE INGREDIENTS IN COACHING?

New findings on the effectiveness of the coaching relationship:

time to think differently about active ingredients?

Abstract

This article critically reviews two recent, large-scale, randomized controlled trials in executive coaching, in order to drive further exploration into the topic of the *coaching relationship* as a predictor of coaching outcome. One of the trials was designed at senior levels in an industrial setting and the other was an experiment with coaching in a business-school context. Each trial demonstrated considerable and significant coaching effectiveness with the coaching relationship ('working alliance') as an important ingredient of effectiveness. The more recent randomized-controlled-trial sample, which was longitudinal, seems to show that we may have to radically change our understanding of the impact of the coaching relationship on coaching effectiveness. Contrary to previous consensus, it seems the working alliance between client and coach is not strongly related to coaching effectiveness. The strength of the working alliance only correlates with a higher effectiveness score from the beginning of the coaching relationship, but it does not significantly correlate with increasing outcomes through further coaching conversations. Some possible explanations for this unexpected and seemingly contradictory finding in the area of 'working alliance' are put forward and critically reviewed.

What's It Mean? Implications for Consulting Psychology

From experience and previous research, a good relationship or "working alliance" between coach and coachee was understood to be an important factor in achieving a demonstrable outcome in executive coaching. This article shows that we cannot yet claim a direct impact of the alliance on the incremental results from session to session.

Introduction

IS IT TIME TO THINK DIFFERENTLY ABOUT ACTIVE INGREDIENTS IN COACHING?

Executive coaching is a tailored form of organisational and leadership development where a leader has a series of contracted and confidential conversations with a coaching psychologist or development expert. It is a form of organisational learning through one-to-one conversations, which facilitates development for an individual. It can be used in a variety of ways, for example, getting past an impasse, reflecting on a critical decision, or drawing out and building on strengths. Undoubtedly, coaching is becoming increasingly important. Firstly, there is a general impression that our rapidly changing businesses demand a higher engagement from the whole person in organisations (i.e., including our personal sensitivity, feeling and intuition). Executive coaching being so tailored to the goals and needs of the leader, seems to be particularly well placed to address this demand. Secondly, there is a growing number of successful implementations, e.g. 'internal coaching', 'performance coaching', 'first-100-days coaching' and 'team coaching'. Thirdly, effectiveness of executive coaching seems to be rather well demonstrated in recent years. Moreover, the research done in this field confirms similar studies in adjacent professions such as mentoring, counselling and psychotherapy. The big question that still remains is *what makes coaching so effective*, i.e., what are its 'active ingredients' (McKenna & Davis, 2009)? If we knew more about such factors of effectiveness, we could adapt training, education and selection of coaches, and model our coaching approaches, interventions and contracts around those ingredients that will have been demonstrated to contribute most to outcome. However, clear consensus about 'active ingredients' has so far been elusive and recent research seems to indicate why this may be the case.

Background: what we know about coaching effectiveness

Coaching effectiveness research has been undertaken since the early nineties (Peterson, 1993) and there are now many studies including large-scale, cross-sectional and randomized-

controlled trials (RCT) demonstrating significant coaching outcomes. Here is an overview of some of the most rigorous outcome studies to date:

Smither, London, Flautt, Vargras & Kucine (2003) designed a large (nonrandomized) control group and conclusions were based on more objective criteria than evaluations by the coachees, namely 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° 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; p. 29).

Evers, Brouwers & Tomic (2006) measured self-efficacy beliefs and outcome expectancies, on each of three dimensions. Their study compared a pre-intervention and post-intervention measurement and 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 $p<0.05$). Again, the intervention itself was short with an average of only four coaching sessions.

Duijts, Kant, Van den Brandt & Swaen (2008) undertook 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. The intervention group self-reported improved health, whilst the findings on most objective measures were not significant with one exception: statistically significant reductions in sickness absence could be demonstrated in this group (2.5 days less off work on average in the year following coaching, than in the intervention group; p<0.01).

Nieminen, Smerek, Kotrba & Denison's (2013) quasi-experimental study followed 469 managers, 227 of whom received 4 or 5 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). However, managers had been allocated to the two groups based on nonrandom selection methods, those in the coached group comprising managers who had recently been promoted and had fewer years of leadership than the non-coached group.

These studies and others have been analyzed in recent systematic (meta-analysis) reviews such as Theeboom, Beersma & van Vianen, 2014; Jones, Woods & Guillaume, 2015; Grover & Furnham, 2016; Athanasopoulou & Dopson, 2018. Most studies show a somewhat lower effect size than in psychotherapy, which are normally found to be higher than 0.7 (Norcross & Goldfield, 2019). This could be due to the fact that coaching sessions are organized at lower frequency and perhaps with less pressure on results.

This article introduces two large-scale randomized controlled trials (RCTs) that have recently been undertaken to gain more clarity on the active ingredients of executive coaching. In the first RCT executive coaching took place at senior levels in a global, industrial setting in

healthcare, and in the other RCT the executive coaching was offered to (post-)graduate students in a business-school context.

Methodology

At the pre-stage, ethical approval was requested and granted by the researchers' universities and organizational authority for the first study was obtained conditional upon anonymity.

Participants

(*Coachee Details – first study*) 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 executive-coaching 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. One hundred eighty of 209 female global senior leaders on the program (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. Total number of coaching sessions were 12 but after six months six monthly sessions had been completed on average.

(*Coachee Details – second study*) Invitations to participate in the study were sent to 3,097 undergraduate students at a UK University Business School to yield a volunteer sample of 226 students (response rate 7%). 105 students were randomly allocated to the experimental group, i.e., to receive coaching, with a further 105 randomly allocated to the 'waiting list' control group. The remaining 16 students were randomly allocated into two 8-strong

experimental and control reserve samples from which to randomly select replacements should any students from the experimental or control groups drop out during the study - hence avoiding leaving budgeted coaching sessions unused. We chose to collect responses to our primary outcome variables at eight time points. This was predicated both by the number of coaching sessions (6), and by the hypothesised non-linear shape of change in our outcomes.

(*Coach Details – first study*) 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. Sixty-six coaches (response rate 61%) completed at least one questionnaire.

(*Coach Details – second study*) 233 accredited coaches, the entire database held by a different UK-based Business School, were invited to take part in the study; that is, to coach one student each for a series of six coaching sessions, held approximately monthly over a period of five months, at a pay rate of $55 per session. Of the 233 coaches invited to the study, 114 (49%) responded affirmatively. However, due to budgetary constraints, only 105 of these were (randomly) selected to receive an invitation to coach a student, with the remaining 9 held in reserve in case any of the chosen coaches dropped out early on in the study.

(*Line Manager Details – first study*) All coachees in the first study had a line manager who was invited to participate in the research. 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

Other than basic demographic variables such as gender, nationality, country of residence and ethnicity, the content of each questionnaire per cohort is summarized in Table 1.

INSERT TABLE 1 ABOUT HERE

1. **Outcome measures**

Coaching effectiveness (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, Duckworth, Birch & Jones, 2013, and De Haan, Grant, Burger & Eriksson, 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.

Goal Attainment was measured using a Goal Attainment Scale (GAS; Spence, 2007), adapted from Grant, Curtayne & Burton (2009), and based on the two goals the students in Study 2 had selected at the beginning. It contains 2 items: the degree of success with a question "Up to today how successful you have been in achieving this goal? (the response coding for all items was 0-100% scale from 0 "no achievement at all" to 100 "total success") and the degree of difficulty with a question "How difficult is this goal?" (the response coding for all items was 1 "very easy" to 7"very difficult"). The overall goal attainment score for a goal is calculated by multiplying the goal's difficulty rating by the degree of success. For those students being coached, coaches used the same scale to give their opinion of their coachee's goal attainment at time points 2 to 7.

Perceived Stress was measured using the 4-item short-form of the Perceived Stress Scale (PSS) of Cohen, Kamarck & Mermelstein (1983), which measures the degree to which a

person evaluates situations in his/her life as stressful. A sample item is "In the last month, how often have you felt that you were unable to control the important things in your life?", and the response coding for all items ranged from 1 "never" to 5 "very often".

2. Independent measures

Mental wellbeing was assessed with help of the 14-item Warwick-Edinburgh Mental wellbeing 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 wellbeing.

Perceived Social Support was measured with help of Zimet's Multidimensional Scale of Perceived Social Support (MSPSS; Zimet, Dahlem, Zimet & Farley, 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 "there is a special person who is around when I am in need".

Hope was measured for all coachees using the Adult Hope Scale (AHS; Snyder *et al.*, 1991). This consisted of 12 items (sample item: "There are lots of ways around any problem"), which separate into two 4-item subscales measuring the underlying dimensions of Agency (i.e. goal-directed energy) and Pathways (i.e. planning to accomplish goals) - and 4 further 'filler' items, which we included in the survey but discarded for measurement purposes. The response coding for each item ranged from 1 "Definitely False" to 8 "Definitely True".

General Self-Efficacy was measured for coach and coachee (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".

Coaching Outcome Expectancy was measured by the 6-item Credibility/Expectancy Scale (Devilly & Borkovec, 2000), adapted for coaching. This scale was given to coachees in an attempt to measure how both believable, convincing, and logical undertaking a hypothetical coaching regime would be to the client prior to it actually being offered. It measures two underlying dimensions, credibility and expectancy, each measured by three items (sample items "At this point, how successfully do you think these coaching sessions will be helping in addressing the issues you want to deal with?"; "At this point, how much do you really feel that coaching sessions will help you to reduce issues you want to address?"). Responses were on a 9-point scale, ranging from "not at all logical" to "very logical" or from "not at all useful" to "very useful".

Resilience was measured with the Brief Resilience Scale (BRS; Smith *et al.*, 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-Coachee Relationship was measured with the Working Alliance Inventory (WAI; 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, in order 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".

Coach Behaviors of the coach were measured from the viewpoint of coachee and coach, using 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.

Personality Characteristics were measured using the Hogan Personality Suite (Hogan & Hogan, 1997) and Myers-Briggs Type Inventory (MBTI; Myers, McCaulley, Quenk & Hammer, 1998). 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.

Data analysis

10

(First study) After computing descriptive statistics, reliability estimates, and intercorrelations for all time points we undertook Confirmatory Factor Analysis (CFA), by testing a configural invariance model across all time periods. Then, on the lookout for active ingredients that have an impact on coaching outcomes and personality, we regressed all variables on a model in accordance with our hypotheses.

(Second study) We undertook a complete scale validation process and with the help of Latent Growth Curve Modelling (LGCM) we modelled change over time in our mediator and outcomes. We tested the hypotheses pertaining to the effect of working alliance on our outcomes separately as this model was only applicable to the subsample of students who had received coaching.

Results: summary of the two experiments

The first experiment is the largest RCT undertaken to date in coaching, in a realistic environment for executive coaching: a global healthcare company, comprising approximately 100,000 employees based in over 120 countries (De Haan, Gray & Bonneywell, 2019). It involved two consecutive groups on a leadership-development program which was exclusively based on coaching and 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' design. The study examines the relative impact of various "common factors", with the help of the largest RCT to date. By asking coachees, coaches and line managers to rate outcomes on the same scales three independent measures of coaching outcomes were obtained.

This RCT found strong support 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, with effect sizes *d* larger than 1, confirming meta-analysis studies such as Jones *et al.* (2015). By using a

randomized control group of coachees in exactly the same circumstances, we can safely assume 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 can be assured that the finding is robust against same-source bias. This study also found support for common factors contributing to this effectiveness, in particular coachee-related factors such as resilience, self-efficacy, perceived social support and mental wellbeing, and also for the working alliance between coach and coachee. Although the study was undertaken within the healthcare industry (just like Grant, Curtayne & Burton., 2009), it seems likely that these findings are globally generalizable over many industries, because the coachees were globally mobile, senior and mostly general managers, not technical healthcare experts.

In the second RCT study (Molyn, De Haan, Stride & Gray, 2020), a slightly larger group of 105 business-school students at a London university were coached by an equal number of qualified coaches, whilst another 105 students formed the control group. Data was collected over eight data points for all participants, enabling the researchers to model the dose-effect curve type change expected over the course of, and following the coaching sessions (see Figure 1). Responses were collected from both students and coaches, in order to again overcome issues of purely self-reported measures. Again, effect sizes d over the 6-month journey were between 0.9 and 1, indicating large effects. Furthermore, in this study significant change was established on a range of parameters: coaching effectiveness, resilience, stress and goal attainment scores. Again, there was unmistakable evidence for common factors contributing to this effectiveness, namely hope, outcome expectations, self-efficacy, perceived social support, and also the working alliance between coach and coachee.

INSERT FIGURE 1 ABOUT HERE

It is reassuring to see two RCTs in executive and workplace coaching find the same effect sizes in very different contexts (senior manager coachees vs. student coachees; global business setting vs. local business-school setting; leadership development programme vs. paid subjects in student context; two very different groups of qualified coaches with a different 'technique' base, etc.). The two RCTs find effectiveness both on 'coaching effectiveness', whilst the first study also measures the impact on leadership personality and the second study measures effects on other variables such as goal achievement, resilience and wellbeing improvement, and stress reduction.

However, there were also important differences. In the second study it was only possible to detect significant effects of hope, perceived social support, psychological wellbeing and working alliance ratings on *the levels* of effectiveness, not on the *change* in effectiveness during the course of coaching. This means it is now harder than before to associate these variables with the effectiveness of coaching per se: psychological wellbeing and working alliance were correlating substantially with the coaching effectiveness from beginning to end (see Table 2 for an illustration with the second timepoint measurement), but they were not correlating with the gains through coaching. We believe that this means that we will have to revise our thinking about coaching effectiveness profoundly.

INSERT TABLE 2 ABOUT HERE

Discussion: we still know little about the active ingredients of coaching

For many years now the research literature in the helping professions has told a remarkably coherent story, entailing an affirmative answer to the question "does psychotherapy/counselling/mentoring/coaching work?" and broad consensus on the second question "what are the factors that contribute to its effectiveness". Experiment after experiment, study after study, whether in coaching or psychotherapy, confirmed the following

argument (see Wampold, 2001, for an important proponent of this 'great psychotherapy debate'): "(1) Yes, psychotherapy is effective, with effect sizes of $d > 0.7$, which means that at least some 80% of people who choose to be coached (or, who attend psychotherapy) are better off than those in the very same circumstances who do not. (2) The factors that make helping conversations effective are 'common factors' to all different techniques and approaches, so the particular ideology, technique or training of the therapist is not relevant, and all schools are equally effective (McKenna & Davis, 2009). This conclusion is often referred to as the 'dodo bird verdict' from Alice in Wonderland. In that tale (Chapter 3) a Dodo bird organised a caucus race and at the end everyone asked it who had won. The Dodo thought long and hard and then it said, "Everybody has won and all must have prizes" (for an illustration in the coaching field that specific 'technique' makes no difference, see the findings of De Haan, Culpin & Curd, 2011). (3) Of these common factors, the most important ones are coachee-related, but the largest one that is determined by the helping conversations themselves, often called the 'best predictor of outcome', is the *working alliance* between therapist and patient (McKenna & Davis, 2009; Martin, Garske & Davis, 2000), or between coach and coachee (Grassman, Scholmerich & Schermuly, 2019). This story may still very well be true. However, based on the two recent experiments summarised above, we believe that there is now a counterexample to argument (3), which may also affect whether one trusts argument (2). Because the research literature is mostly built up through correlational data and not as much through longitudinal studies, we feel this counterexample is worth paying attention to.

Moving beyond simple correlational evidence

The older studies on the impact of the coaching relationship, which have recently been summarised in a meta-analysis (Grassman *et al.*, 2019), show consistently that working alliance (as measured by the Working Alliance Inventory, WAI) is linked to outcome. This is

true for coachee-rated WAI (e.g. Baron & Morin, 2009; Boyce, Jackson & Neal, 2010), but also, to a lesser extent, for coach-rated WAI (De Haan *et al.*, 2016) and even for observer-rated WAI on the basis of video-recordings (Gessnitzer & Kauffeld, 2015), so Working Alliance estimates seem to be powerfully related to outcome. Yet they also seem to be independent between different stakeholders: coach, coachee and observer ratings of WAI do not correlate significantly (De Haan *et al.*, 2016; Gessnitzer & Kauffeld, 2015).

In sum, *even though* WAI correlates with coaching outcome throughout, i.e. at different time points, even at much later time points (Martin, *et al.*, 2000), early WAI *does not correlate* with any *increase* in coaching outcome from the second session onwards (as coach and coachee need to have met and worked together, WAI can only be measured after the first session). This can only be explained as WAI being important in terms of a general readiness for coaching, i.e., for initial or average values of outcome, but not for the impact, the effectiveness of the coaching intervention itself.

There have been other doubts expressed about the use of WAI over the years. There are different instruments and they measure different aspects of the alliance. WAI is now the most commonly used, but most authors find that client- and therapist-versions of this questionnaire often do not correlate with each other (Fenton, Cecero, Nich, Frankforter & Carroll, 2001); even that only the observer version has a good correlation with objective outcomes (Gessnitzer & Kauffeld, 2015); and there is some critique of the 'Bond' scale which tends to correlate somewhat less with outcome (see Table 1 and also De Haan *et al.*, 2016).

There has been mainly correlational research into the impact of the alliance, involving only a single measurement of WAI, and therefore causality is beyond the reach of most studies. One study in psychotherapy that did look into causality by measuring WAI and Outcome four times, found only a small yet significant link between working alliance and outcome (Zilcha-Mano, Dinger, McCarthy & Barber, 2014), just as Molyn *et al.* (2020) did.

Our interpretation of these findings

We believe that WAI has perhaps been misconstrued and misnamed as a relational variable, one that might tell us about the strength of the relationship in the coaching room. Instead, it is perhaps more justified to see it as a measure of a *coachee propensity to relate*, i.e. not a relational variable but a client-related variable. The Working Alliance ratings by a coachee may tell us mostly about the coachee, about how disposed the coachee is, generally, towards a good working relationship, and how easily the coachee thinks s/he engages in a relationship that s/he rates positively (e.g., from the Agreement on Task, Agreement on Goal, and Bond perspectives). This picture is confirmed by the fact that coachee and coach scores of the working alliance only show a very limited correlation (De Haan *et al.*, 2016 and 2019) and by the fact that observer scores behave differently from those of the coachee and coach (Gessnitzer & Kauffeld, 2015). In our view, it is like having observers or coaches estimating the coachee's self-efficacy, or wellbeing: they would struggle to do so in accordance with the coachee's own scores.

All earlier findings (summarized by Grassman *et al.*, 2019) can now be understood in our view to show that coachees who say they 'relate' better also achieve better results in coaching, at least according to their own estimates, occasionally confirmed by their coaches and line managers. We think WAI measures a relatively stable coachee personality 'trait', because it seems rather stable and generalizable over time throughout therapy (Martin *et al.*, 2000; Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz & Gallop, 2011), although it may also be measuring a more changeable aspect of personality, a 'state' (Zilcha-Mano, 2017). The fact that WAI levels do tend to increase during the coaching journey (also in Molyn *et al.*, 2020) confirms that there are movable, 'state'-like aspects present in WAI scores.

However, what is important for coaching outcome research, is that the coachee's relationship scores *do not* seem to drive outcomes: the sense of relating well with your coach gives coachees a higher outcome over-all, but hardly impacts on further change by the sessions themselves. In fact, in the second study it was shown that the coachee's Resilience scores are a much better predictor of outcome, to such an extent that most of the predictive power of other variables were picked up by Resilience which therefore seems to be a good 'mediator variable' for effectiveness (Molyn *et al.*, 2020).

This helps to explain why so little evidence has been found for additional effectiveness to be had from particular 'matching' between coachee and coach (Boyce *et al.*, 2010; Page & De Haan, 2014; Bozer, Baek-Kyoo & Santora, 2015): there are certain traits that give coachees an over-all increase in effectiveness (such as WAI, hope, self-efficacy, etc.) but they are hardly further improved by the sessions themselves or by the coach match.

It appears that coachees rate their experience in coaching 'holistically', or as some authors put it, with a 'halo effect' (Dion, Berscheid & Walster, 1972): the coachee will rate all aspects of the experience as better or worse, in accordance with how useful the general experience was for them and their own optimism about these kinds of experiences (significantly influenced by coachee-based factors such as hope, expectancy, self-efficacy, resilience, mental wellbeing, and now, in our view, working alliance) and will then score all other aspects accordingly. Coachee personality aspects like 'Openness', which are also known to correlate positively with outcome (Stewart *et al.*, 2008), might also be influenced by this halo effect. Part of the effect may be to do with 'cognitive dissonance reduction' (a form of 'placebo'): coachees may score any or all of these variables higher after coaching because they have gone through the whole effort of engaging with coaching and participating in six or more sessions. This may even be true for the coach- and line-manager-reported data, i.e., the controls

17

for 'same source' bias may not have been sufficient if we are only asking coaches, coachees and their line managers.

This interpretation would explain why coachee ratings of effectiveness tend to correlate with self-scored working alliance, but also with *any* coaching technique as scored by the coachee on e.g. the CBQ (see De Haan *et al.*, 2011 and 2019). It also explains the high degree of 'reverse causality' which is often found, where coaching is not only impacted by but in turn improves aspects such as self-efficacy, working alliance, and resilience (see e.g. Evers *et al.*, 2006; and Baron & Morin, 2009). In our second sample WAI was significantly related to future CE (for most time periods), but there was no reverse causation of WAI.

On the plus side, each of the studies does demonstrate a particular aspect that might make executive coaching a very helpful intervention for senior leaders: the second study shows that *resilience* is a particular driver of outcome, and resilience does tend to be a characteristic of senior executives. The first study, moreover, showed that coaching has a significant beneficial effect on personality derailers in leadership, something that is highly relevant at top levels of organisational leadership (De Haan & Kasozi, 2013).

Limitations of this research note

This contribution was made on the basis of only two relevant samples, two large-scale RCTs that have recently been undertaken. Even though they were embedded in the previous literature, they have here been reviewed from the perspective of the coaching relationship only. Although some of our argument is corroborated by recent findings in psychotherapy research, as indicated, it is still very early to say if these results are generalizable. In order to gain more certainty for or against our novel argument, we would need many more independent longitudinal trials studying executive coaching outcomes (even in the larger research base of psychotherapy there is still some controversy as there are also studies that do indicate a

sustained impact of WAI on effectiveness, e.g. Falkenström, Granström & Holmqvist, 2013) Furthermore, there were clear constraints on each sample of coaching clients: one only consisting of female leaders in a very particular industry, and the other consisting of business-school students only. However, as the quoted articles also make clear, we have found no evidence of biases in comparing these groups to each other and to other samples in the research literature.

Future research: time to think differently about active ingredients

In our opinion there is still a lot more work to do to discover what exactly makes coaching effective. Most of the factors found to date are coachee-related and not really related to anything that goes on in the sessions themselves. We believe it is important to start thinking differently in terms of outcomes. Specific techniques or approaches, specific common factors, are very difficult to demonstrate and may not be relevant. Although we should of course for the time being keep an open mind and test other self-scored relationship measures as well, particularly those that seem to refer more strongly to the actual coaching environment that is shared by both parties. In this regard there are some preliminary studies that demonstrate alternative ways to measure a positive coaching relationship e.g. through fMRI, salivary cortisol samples and coachee positive-emotion word counts: Jack, Boyatzis, Khawaja, Passarelli & Leckie (2013) and Howard (2015).

It seems however that potential active ingredients in coaching are more general than scholars have suggested so far: they seem to point to a general optimism and stamina on the client's side (hope, self-efficacy, resilience) including generally positive feelings about techniques (CBQ) and the relationship (WAI). We will therefore have to look at more generalized and objective outcome measures, more longitudinal trials, and also at 'holistic' variables that can capture the whole experience in an integrated way, such as (1) around

technique and behaviour, (2) around personality of coachee and coach, (3) around the 'in-between' or coaching relationship. The latter should perhaps focus more on the atmosphere in the room and the critical moments in sessions and should perhaps first be investigated through scoring by observers who are more likely to be objective about what they see in the conversation (following e.g. Gessnitzer & Kauffeld, 2015).

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.

Baron, L. & Morin, L. (2009). The coach-coachee relationship in executive coaching: A field study. *Human Resource Development Quarterly, 20*, 85-106.

Bozer, G., Baek-Kyoo, J. & Santora, J.C. (2015) Executive coaching: does coach-coachee matching based on similarity really matter? *Consulting Psychology Journal: Theory and Practice, 67*(3): 218-233.

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.

Cohen, S., Kamarck, T. & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), 385-396.

Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S. & Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology, 79*, 267-278.

De Haan, E. Culpin, V. & Curd, J. (2011). Executive coaching in practice: What determines helpfulness for clients of coaching? *Personnel Review, 40,* 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: Practice and Research*, 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.

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

De Haan, E. & Kasozi, A. (2014). *The leadership shadow: how to recognise and avoid derailment, hubris and overdrive.* London: Kogan Page.

De Haan, E. & Nilsson, V. (2017). Evaluating coaching behavior in managers, consultants and coaches: A model, questionnaire, and initial findings. *Consulting Psychology Journal: Practice and Research,* 69(4): 315-333.

Devilly, G.J. & Borkovec, T.D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry,* 31, 73-86.

Dion, K., Berscheid, E. & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285–290.

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 Environmental Medicine, 50*.7, 765-776.

Evers, W. J. G., Brouwers, A. & Tomic, W. (2006). A quasi-experimental study on management coaching effectiveness. *Consulting Psychology Journal: Practice and Researcher, 58*(3), 174-182.

Falkenström, F., Granström, F. & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*, 317–328.

Fenton, L. R., Cecero, J.J., Nich, C., Frankforter, T. L. & Carroll, K. M. (2001). Perspective is everything: The predictive validity of six working alliance instruments. *Journal of Psychotherapy Practice and Research, 10*(4), 262–268.

Gessnitzer, S. & Kauffeld, S. (2015). The working alliance in coaching: Why behavior is the key to success. *Journal of Applied Behavioral Science, 51*(2), 177–197.

Grant, A.M., Curtayne, L. & Burton, G. (2009). Executive coaching enhances goal attainment, resilience and workplace wellbeing: a randomised controlled study. *The Journal of Positive Psychology 4(5)*, 396-407.

Grassman, C., Scholmerich, F. & Schermuly, C.C. (2019). The relationship between working alliance and client outcomes in coaching: A meta-analysis. *Human Relations*, 1-24.

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.

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.

Howard, A. R. (2015). Coaching to vision versus coaching to improvement needs: a preliminary investigation on the differential impacts of fostering positive and negative emotion during real time executive coaching sessions. *Frontiers in psychology 6*, 455.

Jack, A.I., Boyatzis, R.E., Khawaja, M.S., Passarelli, A.M. & Leckie, R.L. (2013). Visioning in the brain: An fMRI study of inspirational coaching and mentoring. *Social Neuroscience 8*(4), 369-384.

Jones, R.J. and Woods, S.A. & Guillaume, Y. (2014) A Meta-Analysis of the Effectiveness of Executive Coaching at Improving Work-Based Performance and Moderators of Coaching Effectiveness. In *British Psychological Society Annual Division of Occupational Psychology Conference, 06/01/14 - 08/01/14*, Brighton.

Martin, D.J., Garske, J.P. & Davis, M.K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. In: *Journal of Consulting and Clinical Psychology 68*, 438-450.

McKenna, D.D. & Davis, S.L. (2009). Hidden in plain sight: The active ingredients of executive coaching. *Industrial and Organisational Psychology*, 2(3), 244-260.

Molyn, J., De Haan, E., Stride, C. & Gray, D.E. (2020). The Contribution of Common Factors to Coaching Effectiveness: Lessons from Psychotherapy Outcome Research. *Journal of Applied Psychology,* to be submitted.

Myers, I. B., McCaulley, M. H., Quenk, N. L. & Hammer, A. L. (1998). *MBTI manual*. Palo Alto, CA: Consulting Psychologists Press.

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.

Norcross, J.C. & Goldfried, M.R. (2019; Eds.). *Handbook of psychotherapy integration*. New York, NY, US: Oxford University Press.

Page, N. & De Haan, E. (2014). Does coaching work? ...and if so, how? *The Psychologist, 27*.8, 582-586.

Peterson, D.B. (1993). Measuring change: *A psychometric approach to evaluating individual coaching outcomes.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Schwarzer, R. & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, and M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, NFER-NELSON.

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, 194-200.

Smither, J. W., London, M., Flautt, R., Vargras, 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.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., Gibb, J., Langelle, C. & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60,* 570-585.

Spence, G.B. (2007). GAS powered coaching: Goal Attainment Scaling and its use in coaching research and practice. *International Coaching Psychology Review*, 2(2), 155–167.

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), 32-43.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S. & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental wellbeing scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, 5(1), 63-76.

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

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods and findings.* Mahwah, NJ: Lawrence Erlbaum.

Zilcha-Mano, S., Dinger, U., McCarthy, K. S. & Barber, J. P. (2014). Does alliance predict symptoms throughout treatment, or is it the other way around? *Journal of Consulting and Clinical Psychology, 82*, 931-935.

Zilcha-Mano, S. (2017). Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. *Am. Psychol. 72*:311–25.

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.

FIGURES AND TABLES

FIGURE 1. The first dose-effect curve in executive coaching, confirming clear benefits throughout the six coaching sessions and a diminishing return per session. Time T1 is before the first session, time T7 is shortly after the 6th session, and time T8 is a 3-month follow up, which explains why the curve bends back.
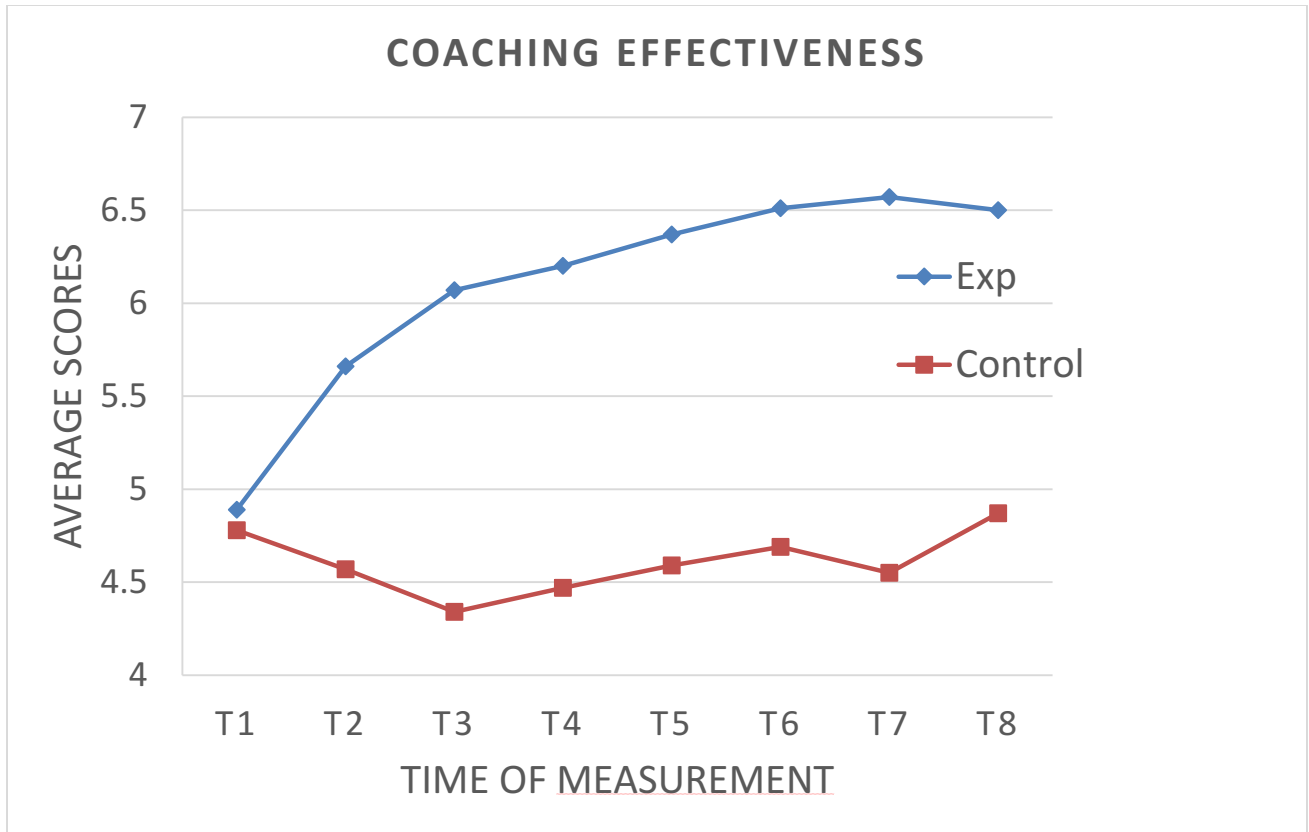
TABLE 1: **Measurement Table Showing Psychometrics Taken by Cohort (Coachee, Coach, Line Manager), and Time Points (T1-T3 for Study 1 and T1-T8 for Study 2). A '1' in the table indicates Study 1 and a '2' indicates Study 2.**

Study Variables:	Coachee T1	Coachee T2-T8	Coach T1	Coach T2-T8	Line manager T1	Line manager T2-T3
Coaching Effectiveness (CE)	1+2	1+2		1+2	1	1
Goal Attainment	2	2	2	2		
Perceived Stress (PSS)	2	2				
Psychological wellbeing scale (WEMWBS)	1+2	1+2				
Perceived Social Support (MSPSS)	1+2	1				
Hope (AHS)	2					
General Self-Efficacy	1+2		1			
Coaching Outcome Expectancy	1+2	1	1	1	1	1
The Brief Resilience Scale (BRS)	1+2	2				
Coach-Coachee Relationship (WAI)		1+2		1+2		
Coaching Behaviors Questionnaire (CBQ)		1		1		
Hogan "bright side" (HPI)	1	1				
Hogan "dark side" derailers (HDS)	1	1				
Hogan "motives, values, preferences" (MVPI)	1	1				
MBTI Personality	1		1			

TABLE 2. Correlations between the Working Alliance Inventory (WAI) and Psychological Wellbeing (WEMWBS) as measured by the coachee, and the coachee's Coaching Effectiveness scores at all time measurements. T1 is before the first session, T7 is after the sixth session, and T8 is a three-month follow-up.

	CE T2 (n = 111)	CE T3 (n = 99)	CE T4 (n = 96)	CE T5 (n = 97)	CE T6 (n = 84)	CE T7 (n = 76)	CE T8 (n = 93)
WAI – Task T2	.56***	.27**	.28**	.59***	.44***	.28*	.43***
WAI – Bond T2	.47***	.27**	.27**	.44***	.30**	.19	.34**
WAI – Goal T2	.51***	.35***	.30**	.54***	.39***	.27*	.35**
WEMWBS T2	.28**	.10	.02	.21*	.23*	.13	.14