Evaluating a Coaching and Mentoring Programme: challenges and solutions

Tatiana Bachkirova, Linet Arthur and Emma Reading

Abstract

Objectives:

This paper describes an independently conducted research study to develop appropriate measures and evaluate the coaching/mentoring programme that the London Deanery had been running for over five years. It also aims to explore specific challenges in the evaluation of a large-scale coaching programme and to suggest new solutions.

Design:

The challenges to evaluation included the need to use established but also context-relevant measures and the need for a rigorous but also pragmatic design that took into account a number of practical constraints. Overall it was a mixed method research design consisting of a within-subject quantitative study with support of a qualitative grounded theory methodology conducted in parallel.

Method:

The selected measures for the quantitative part of the study included employee engagement, self-efficacy and self-compassion. An additional questionnaire SWRQ (Specific Work-Related Questionnaire) was developed as the result of a qualitative investigation with stakeholder representatives. It included a self-estimation by the coached clients of the extent to which they could attribute each change to the coaching received rather than any other factor. The qualitative part of the study included interviews with stakeholders and the analysis of responses to an open question in the SWRQ.

Results:

One hundred and twenty (78%) of matched responses pre- and post-coaching were analyzed and 7 stakeholders interviewed. The results of the quantitative and qualitative analysis show improvement in all chosen scales. The analysis also shows that coaching was a major contributor to these changes.

Conclusions:

The paper argues for the development of additional methods in outcome research on coaching programmes that are aligned with the main principles and philosophy of coaching as a practice.

Introduction

For the last two decades we have been witnessing unprecedented growth of the coaching industry. Many organizations invest in coaching programmes. It seems, however, that the pace of growing evidence of the added value of coaching that should come from research, does not yet match the speed of the implementation of coaching programmes. Consequently many organisations aim to gather their own evidence about the effectiveness of these programmes to justify the return on investment. The London Deanery is one such organisation.

The London Deanery established a Coaching and Mentoring service for doctors and dentists in London in 2008. The coaches were trained by an established leadership coaching provider and their performance was assessed at the end of the training. The outcome of the service was measured by individual feedback from the service users. However, it was viewed that although this provided some data on how the service was performing, it was not sufficient to identify any performance changes in the recipients. The service was publicly funded thus it was important that it should be properly evaluated to ensure value for money. Preliminary work looking at the literature evaluating the benefits of coaching and mentoring did not reveal an established methodology for conducting such a review. The Oxford Brookes team of researchers won a bidding process for the research based on their proposals for developing novel methodologies for the evaluation of the service. The aim of the study was to establish whether the measures selected could identify changes in the performance and attitudes of doctors undergoing the coaching intervention,

since ultimately the purpose of the programme is to improve the effectiveness of doctors and dentists for the benefit of the patient.

It appears that the London Deanery task is not dissimilar to questions asked by many HRD practitioners. However, as Lawrence and Whyte (2014) recently argued, these practitioners "have not yet collectively identified a satisfactory approach to evaluating the efficacy of coaching" (2014: 6). This is not surprising because the problem of measuring the impact of organizational intervention is not new: it has been actively discussed since Kirkpatrick's (1977) methodology for evaluating training programmes (e.g. Ely et al, 2010) but it is still debated, particularly in relation to the evaluation of training (Passmore and Joao Velez, 2014) Kirkpatrick admitted that it is extremely difficult, if not impossible, to evaluate certain programmes in terms of the results (1977) because of the numerous factors influencing the outcomes. Others authors continue echoing this conclusion (Ely et al, 2010; de Haan & Duckworth, 2012). A financial addition to Kirpartick's methodology – ROI (Return on Investment) has been particularly critiqued in relation to coaching with recent conclusions that are not dissimilar to Kirpatrick's premise (De Meuse, 2009; Grant, 2012, 2013; de Haan & Duckworth, 2012; Passmore & Fillery-Travis, 2011; Theeboom et al, 2013).

The evaluation of coaching programmes is considered to be even more difficult than the evaluation of training programmes (Ely et al, 2010; Grant, 2012, de Haan & Duckworth, 2012). The problems are exacerbated by a number of factors, such as the diversity of outcomes of coaching comparing to the relatively fixed expectations of training; the highly individual approach of the coach, which prevents more explicit knowledge of the process; the confidentiality that surrounds specific details of the goals and consequent outcomes.

On the other hand, as coaching programmes tend to be expensive there appears to be a stronger need to justify such expenditure, particularly in the public sector. Therefore, in spite of the additional costs involved in undertaking a full-scale research study on assessing the effectiveness of a coaching programme, the London Deanery chose this option. However, such evaluations face similar issues and obstacles as large-scale outcome research projects on organizational interventions (Theeboom et al, 2013; De Meuse, et al,

2009; Passmore & Fillery-Travis, 2011), particularly the use of established paradigms. If traditional positivist methodologies are considered as the gold standard of evaluation, there is a danger that the complexity of coaching interventions may be overlooked (de Haan & Duckworth; Ely et al, 2010). This paper explores how some specific challenges of the evaluation of a large-scale coaching programme were addressed and will suggest a methodology that is more in line with the philosophy of coaching. This will be discussed together with the results of the actual evaluation.

Literature review

While recognizing that there are wider debates on the evaluation of organizational interventions (for example, Passmore & Joao Velez, 2014), this literature review is only focused on the evaluation of effectiveness of coaching programmes. Typically this literature addresses three main themes: a) issues of evaluation in principle depending on the main stakeholders; b) most acceptable methodologies of evaluation and c) specific measures of evaluations.

In relation to the *general issues of evaluation* Grant (2013) suggests that we need to start with a question: "who is interested in evaluation – and why" (2013: 15). The first group concerned with this question is the coaches. On the one hand they wish coaching to be seen as effective for marketing purposes; on the other hand, they are interested in improving their practice. Purchasers of coaching ask the question of whether coaching works because they want to know if coaching is cost-effective. Both coaches and purchasers seem to have vested interests in the results of evaluation. In comparison, researchers and academics are well placed to explore the effectiveness of coaching using rigorous research methods and are interested in developing evidence-based practice for coaching (Fillery-Travis & Lane, 2006; Briner, 2012; Passmore & Fillery-Travis, 2011). However, they are probably more than others aware of many difficulties in applying

scientifically respected methods to researching coaching practice (Ellam-Dyson, 2012; Drake, 2009; Ely et al, 2010; Grant, 2012).

One of the main problems is that many of these research methods require significant oversimplification of the nature of practice. Coaching is a complex intervention influenced by the interplay of different factors such as the client's attitude, coach's skill, coach/client relationship, all of which are subject to complex dynamics affected by the contextual issues (de Haan & Duckworth, 2012; Ely et al, 2010). In addition if coaching is sponsored by an organisation it is difficult to establish who the main provider of information about the effectiveness of coaching should be: the client, the coach, the purchaser of the service or those on the receiving end of the changes that are made by the client. In terms of more specific issues various authors also question a typical assumption that coaches from different backgrounds, training and styles deliver the same type of coaching and whether it is possible to treat coaching as a homogeneous intervention, allowing general conclusions to be drawn about its effectiveness (Theeboom, 2013).

In terms of *acceptable methodologies of evaluation* the literature confirms again multiple issues with different methodologies. It is accepted that there are certain advantages and disadvantages in each methodology. Grant (2013) differentiates as rigorous three types of outcome studies with an indication of potential issues associated with them:

- Case studies that can provide valuable descriptive data but do not allow generalized evaluations or the comparison of results between different coaching interventions.
- Within-subject outcome research which allows comparison of the impact of coaching on a group of individuals. The group is assessed before and after the coaching interventions. This is the most commonly used study design in the literature and can provide valuable quantitative data of change, but causation cannot be attributed only to coaching.
- Between-subject and Randomized Controlled Studies, which are considered by some to be the 'gold standard' particularly in medical research. Although they can

5

measure change and relate it to the intervention, the utility of these designs for studying coaching is contested (de Haan & Duckworth, 2012; Greif, 2009; Passmore & Fillery-Travis, 2011) due to problems with delineating a control group, maintaining the 'blind' condition and constructing 'placebo' interventions. Uses of self-coaching, peer-coaching or 'waiting list' are considered practical issues in terms of implementation in relation to coaching studies (Hicks, 1998; Franklin & Doran, 2009; Greif, 2009; Williams, 2010).

The traditional research literature on evaluations typically associated with a positivist paradigm, focuses on searching for general relationships between a small number of discrete variables across wide varieties of context. However, these contexts, from a constructionist's point of view, have a large impact upon these relationships (Fishman, 1999: 235). Without consideration of context the findings of such studies may lead to conclusions that are so generic that their practical value becomes questionable (Orlinsky, et al, 1994; Grief, 2007; de Haan & Duckworth, 2012).

It is not surprising that some research communities resist the idea that only one notion of research is recognised as science: the one identified with modernistic positivism. It has been argued that there are other meanings of science, for example, as disciplined, critical, reflective thought that compares and contrasts evidence, arguing for alternative interpretations or explanations of a particular phenomenon (Fishman, 1999; Cronin & Klimoski, 2011). In coaching research Grant (2013) argues, "an evidence base *per se* does not purport to prove that any specific intervention is guaranteed to be effective, nor does it require that a double-blind, randomized, controlled trial is held as being inevitably and objectively better than a qualitative case study approach" (2013: 3). This means that the evaluations of coaching could be approached from different research paradigms (pragmatism, contextualism, interpretivism) and may benefit from mixed designs. For example it can include retrospective questionnaires validated by traditional positivist procedures (Passmore, 2008), but also include new instruments that were developed with

considerations of factors such as the type of coaching, the organizational level of the coachee, the specific objectives and context of each coaching engagement (De Meuse et al, 2009; Ely et al, 2010).

The theme of *specific measures* that could be used for evaluation of coaching is not an easy one either. According to Fillery-Travis and Lane (2006) before we can ask whether coaching works we must ask why it is being used. A fundamental difficulty of coaching outcome research is the extreme heterogeneity of issues, problems and goals, which can be picked out as themes in different coaching interventions. This could be compared with therapy where it is possible to offer general indicators of the quality of service such as subjective wellbeing, symptom reduction and life functioning (e.g. Mental Health Index, Howard et al, 1996). In coaching, however, it is difficult to identify the outcome measures applicable to the whole range of coaching interventions (Greif, 2007, p. 224). Grant (2013), providing many examples from the vast range of issues addressed in coaching, concludes that there is an almost endless list of applications. The majority of these outcomes are difficult to quantify. This is why sometimes the target outcomes are selected because they can be measured, rather than because they are appropriate for individual clients or reflect the nature of coaching (Easton & Van Laar, 2013).

Often practitioners create a battery of measures, which might reflect the context of the study, their priorities and those of their client or organisation commissioning the evaluation. A combination of measures or indicators can sometimes help to avoid oversimplification with an intention to work towards meeting a particular target that is measured (Easton & Van Laar, 2013). Greif (2007), for example, proposes general measures (degree of goal attainment and client satisfaction with coaching) and specific measures such as particular social competences; performance improvement and self-regulation. The choice of these measures has to be justified by the theories tested in the independent research or by the practical needs of the organisation commissioning the evaluation.

Overall, there is recognition in the literature that outcome research and evaluation of effectiveness of coaching face significant challenges. Therefore there is a need for pragmatic and creative approaches to this task, which could assure rigour as the result of competent inquiry.

Methodology of the project

This project was designed as a pragmatic inquiry requiring mixed methods (Tashakkori & Teddli, 2010) with a large proportion of the data, as requested by the client, of a quantitative nature using qualitative data to construct a questionnaire and further inform the results.

The quantitative element of the study aimed to establish whether the coaching and mentoring provided by London Deanery practitioners had an impact on clients by comparing their scores from Time 1 (pre-coaching) and Time 2 (post-coaching) online measures. In this project the London Deanery was interested in two variables, which are theoretically related to the individual change process and were considered by them as relevant for the situation: employee engagement and self-esteem. To give a fair representation of the aspect of self-esteem the researchers suggested two measures: self-efficacy and self-compassion.

Selected measures

Schaufeli & Bakker (2003) describe *Employee Engagement* as a positive, fulfilling, work-related state of mind that is characterized by vigour, dedication and absorption. Research over the past ten years has shown the importance of this concept in relation to understanding key organisational outcomes, such as low turnover (Schaufeli & Bakker 2004), high organisational commitment (Demerouti, Bakker, de Jonge, Janssen & Scaufeli, 2001), and customer-rated employee performance (Salanova, Agut & Peiro, 2005). This has a potential to provide an indirect relationship between the effect of coaching and 'customer-rated' employee performance as well as reduced turnover and increased organisational

8

commitment. Employee engagement is also negatively related to burnout, which was particularly important to the London Deanery

The Oxford Brookes University team used a typical instrument for measuring employee engagement, the Utrecht Work Engagement Scale (Schaufeli & Bakker, 2003). This scale consists of 17 items, six of which measure Vigour, six measure Absorption and five measure Dedication. Vigour is characterized by high levels of energy and mental resilience while working, the willingness to invest efforts in one's work and persistence even in the face of difficulties. Dedication refers to being strongly involved in one's work and experiencing a sense of significance, enthusiasm, inspiration, pride and challenge. Absorption is characterized by being fully concentrated and happily engrossed in one's work, whereby time passes quickly and one has difficulties with detaching oneself from work (Schaufeli & Bakker, 2003).

The scale consists of 7 points from 0 = Never had this feeling, 1 = almost never, a few times a year or less, 2 = rarely, once a month or less, 3 = sometimes, a few times a month, 4 = often, once a week, 5 = very often, a few times a week, to 6 = always, every day. The mean scale score of the 3 subscales is computed by adding the scores on the particular scale and dividing the sum by the number of items of the subscales involved. A similar procedure is followed for the total score.

Perceived *Self-Efficacy* refers to beliefs about one's competence to deal with challenging encounters and the "belief in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p.3). It is clear why this concept is related to beneficial coaching and mentoring outcomes. There is now a large body of research that supports a relationship between measures of perceived self-efficacy and performance (Stajkovic & Luthans, 1998).

Self-Efficacy was measured using the Generalized Self-Efficacy Scale, GSE (Schwarzer & Jerusalem, 1995). Cross-cultural research has been carried out which

confirms the validity of this scale, showing consistent evidence of associations between perceived self-efficacy and other psychological constructs (e.g. health behaviours, wellbeing, social cognitive variables and coping strategies (Luszczynska, Scolz & Schwarzer, 2005).

The 10 items are scored using a 4 point scale: 1= not at all true, 2=hardly true, 3=moderately true and 4 =exactly true. Schwarzer and Jerusalem (1995) found that the internal consistency varied across cultures but ranged from .78 to .91 and concluded that it was very satisfactory, considering that the scale only has 10 items. Scherbaum et al (2006) used Item Response Theory to test the GSE Scale and found that it works best for individuals with average or below average levels of GSE. The GSE Scale is less precise at above average levels of GSE.

As low self-esteem has been frequently associated with negative social comparisons and internalized self-judgments, *Self-Compassion* (Neff, 2009) was introduced as an individual measure that is also a predictor of the ability to cope effectively with adversity and good mental health. Self-compassion comprises being kind and understanding towards oneself when experiencing pain or failure as opposed to being harsh and self-critical. Research studies have consistently linked self-compassion to reduced fear of failure, enhanced perceived competence and emotionally-focused coping strategies, suggesting that this indicator is a promising one for coaching (Neff, 2009; Neff & Vonk, 2009; Neff & Lamb, 2009).

Self-compassion has three basic components: 1) extending kindness and understanding to oneself; 2) seeing one's experiences as part of the larger human experience rather than as separating and isolating, and 3) holding one's painful thoughts and feelings in balanced awareness and not over-identifying with them. (Baumesieter, Bushman & Campbell, 2000, as reported by Neff, 2003).

Self-compassion is measured with the Self-compassion Scale (SCS, Neff, 2003). The 12-item scale was used with items 2 & 6 for self-kindness, items 11 & 12 for self-judgement, items 5 & 10 for common humanity, items 4 & 8 for isolation, items 3 & 7 for

mindfulness and items 1 & 9 for over-identification. Subscale scores are computed by calculating the mean of subscale item responses. To compute a total self-compassion score, reverse score the negative subscale items - self-judgement, isolation and over-identification (i.e. 1=5, 2=4, 3=4, 4=2, 5=1) - then compute a total mean. Neff (2003) found that internal consistency was above the acceptable: .75 level for overall and subscales. Test/re-test reliability of overall scale plus subscales was also found to be acceptable (.93 overall).

A bespoke questionnaire

We made the decision to develop another instrument for this evaluation for two main reasons. The first was responding to the need of the client-organisation to address the more specific contextual relationship between the coaching service provided to London Deanery clients and noticeable behavioural and attitudinal changes that might be linked to their work performance and, consequently, patient care. The second was recognizing the importance of being creative when facing various issues associated with measurement. In this case our intention was to capture not just the static estimation of particular aspects of the clients' working lives but directly addressing the degree of changes that happened in relation to these aspects by the end of the coaching process.

To create this measure we interviewed appropriate stakeholders: three users of service (two consultants and one GP), two coach/mentors and two matchers, those referring clients to this service and identifying a suitable coach. A Grounded Theory approach (Strauss and Corbin, 1990) was used as the main methodology for analysis. The following themes that emerged as the result of the qualitative analysis were particularly useful in the development of the bespoke questionnaire:

a) Impact on patients

- Improved interactions with patients

- Improved feedback from patients

- Use of coaching/mentoring techniques with patients

- Changes in patients' behaviour, such as reduced dependency, better use of doctors' time.

b) Impact on colleagues

- Improved interactions and communication with colleagues

- Use of coaching/mentoring techniques with colleagues.

c) Impact on self

- Improved confidence

- Better time management at work, leading to an improved work-life balance

- Improved capacity to solve problems and make decisions, including career decisions

- Better relationships with family members

- Made them decide to stay within the profession after seriously considering leaving the NHS.

These questions allowed generation of contextually meaningful data for this evaluation. Although this added an important element for measuring the potential impact of coaching we had to acknowledge that in this type of study, without a control group, it was impossible to claim that changes happening to coached clients were the results of coaching rather than any other influences or combination of influences. In order to minimize this limitation we added another question to our bespoke questionnaire in which clients themselves could indicate to what degree coaching contributed to each identified change. Although this indication is a self-estimation we believe that the well-educated and self-aware clients in this study were conscious agents of their life situation and therefore had sufficient insight into the relationship between various influences in their lives. We believed in their unique position to isolate the role of the service they received from the complex array of other factors in their

life and to be completely open about this under the conditions of this particular study when there was no reason for them to misrepresent their responses.

This questionnaire was piloted internally to check the questions and appropriate adjustments were made following feedback. The final questionnaire SWRQ (Specific Work-Related Questionnaire) became part of the Time 2 questionnaires.

To summarize, the Time 1 questionnaire consisted of demographic questions and three scales measuring Employee Engagement, Self-Efficacy and Self-Compassion. The demographic questions were developed to capture the respondents' age group, sex, ethnic origin, whether trained inside or outside the UK and career level. The Time 2 questionnaire included the above three scales of the Time 1 questionnaire and the SWRQ scale that aimed to identify changes in aspects of working life of the client and degree to which the client attributed each change to coaching. All the questionnaires asked for the unique registration number allocated on application in order to pair the Time 1 and Time 2 responses for each individual, whilst maintaining anonymity.

Research process

Once the potential participants of the evaluation research applied for the coaching programme online they were informed about the evaluation study and were given an option to opt out if they did not wish to take part. Once accepted, clients were sent their registration number (CLT number) and then an online link to the Time 1 Survey on Surveymonkey. The next stage was the normal process of the coaching programme as provided by the London Deanery. The participants were rung by one of a small team of matchers, all trained Deanery Coaches. A structured conversation was held with the client checking their reasons for seeking coaching, their understanding of the process and practical requirements such as venue and time. The participant was then sent an email with the description of three coaches attached for them to identify their preferred coach. Clients were offered coaches

outside their specialty and outside their place of work to ensure externality to the coaching process. The coaching intervention consisted of four sessions of 1-1.5 hours taken over a period of six months. When the coaching was completed participants were sent a link with an invitation to complete the Time 2 questionnaires. The research was conducted with consideration of good practice and strict ethical guidelines.

Results

Overall there were 189 Time 1 responses and 137 Time 2 responses. After matching responses and taking out responses where the clients had not completed the minimum number of sessions, there was a total of 120 matched Time 1 & Time 2 responses. Therefore, the final response rate was 78%.

The demographic data show that 48.3% of respondents were aged between 30-39, 20.8% were aged between 20-29, 23.3% were aged between 40-49 and 7.5% were aged between 50-59. There were no respondents aged over 60 years old. The majority of respondents were female (66.7%). Nearly all respondents were trained within the UK (92%). The majority of respondents (46.7%) were less than 2 years Post Qualification with a further 25.8% more than 2 years Post Qualification and 18.3% Foundation Trainees. The majority of respondents were White British followed by 18.3% who were Asian or Asian British: Indian.

The results of the three selected established measures (table 1 and 2) indicate that clients benefited from the programme in relation to each of them.

	Employee Engagement Time 1	Employee Engagement Time 2	Self-Efficacy Time 1	Self-Efficacy Time 2	Self-Compassion Time 1	Self-Compassion Time 2
Mean/Standard Deviation	4.13 (0.78)	4.37 (0.71)	2.99(0.39)	3.17(0.39)	2.98 (0.61)	3.22 (0.63)
Median/Range of scores	4.2(4.20)	4.4(3.94)	3(1.9)	3.1(1.9)	2.92(3.17)	3.25(2.84)

Minimum score	1.60	1.93	2	2.1	1.25	1.58
Maximum score	5.80	5.87	3.9	4	4.42	4.42

Table 1   Descriptive data from Time 1 and Time 2 questionnaires

Table 1 shows the descriptive data for both Time 1 and Time 2 for Employee Engagement, Self-Efficacy and Self-Compassion. The first line of data looks at the means and it is clear that all Time 2 means (average) are higher than the Time 1 means (average).

	UWES Manual Sample	London Deanery Sample
Number	655	120
Mean	3.10	4.13
Coding	"At least a couple of times a month"	"At least once a week"
Nationality	Dutch & Finnish	English
Background	Completed career counselling questionnaire	Applied to coaching programme

Table 2   Comparing UWES sample and London Deanery sample

However, before exploring whether these differences are statistically significant, it is important to consider levels of Employee Engagement, Self-Efficacy and Self-Compassion before the coaching began. Table 2 describes the differences in results of the sample of doctors that was used in the UWES Manual. These results suggest that the clients in this study had higher levels of employee engagement before they started the coaching than the sample from the UWES manual. What is also important to point out is the minimum and maximum scores and the resulting range of scores. Whilst the average employee engagement levels are reasonably high there is a wide range of scores, with the lowest being 1.60 (which equates to " at least once a year") and the highest being 5.80 (which

equates to "a couple of times a week or daily"). At Time 2 the range of scores is reduced, as is the Standard deviation, which measures dispersion around the average value.

Although Self-Efficacy scores had the largest effect size (Table 3) the ranges of scores and standard deviation stayed nearly the same. Like Employee Engagement, scores for Self-Compassion showed decreases in range of scores.

Scale	Standard Deviation	t	Df	Sig. (2-tailed)
Employee Engagement	0.66	3.968	119	0.000
Self-Efficacy	0.36	5.423	119	0.000
Self-Compassion	0.47	5.586	119	0.000

Table 3   Paired sample t-tests to measure whether Time 2 means are higher than Time 1 scores

The results of the paired sample t-tests in Table 3 show a positive impact on mean scores of Employee Engagement, Self-Efficacy and Self-Compassion (at the .01 level) with a highly significant effect for mean scores on all three scales. Effect sizes were calculated based on Cohen's calculations for paired sample t-tests. Effect size for Employee Engagement is .32, effect size for Self-Efficacy is 0.45 and effect size for Self-Compassion is .38. This shows that the effect sizes vary from between small and medium (Employee Engagement and Self-Compassion and medium (Self-Efficacy). It is also important to highlight that there is evidence that the UWES is better at measuring lower compared to high levels of employee engagement. Therefore it is possible that the coaching had more of an impact on coaches than these effect sizes suggest. The General Self-Efficacy scale is also better at measuring lower scores than higher ones.

This means that all three measures selected for their capacity to illustrate meaningful changes in the clients as the result of their coaching, confirm that these changes were

significant. The clients reported higher levels of employee engagement, self-efficacy and self-compassion after being coached in comparison to the levels of these aspects in their lives before they engaged with the coaching programme. The results were not driven by any particular subgroup and benefit was seen across subgroups in race, gender, stage of career and age.

Another type of analysis was made available by using the SWRQ (specific work-related questionnaire developed for this study). This questionnaire was designed to explore the changes that are perceived by the clients in relation to their work. The results of this analysis are shown in the figures 1 and 2 and in the following table 4.
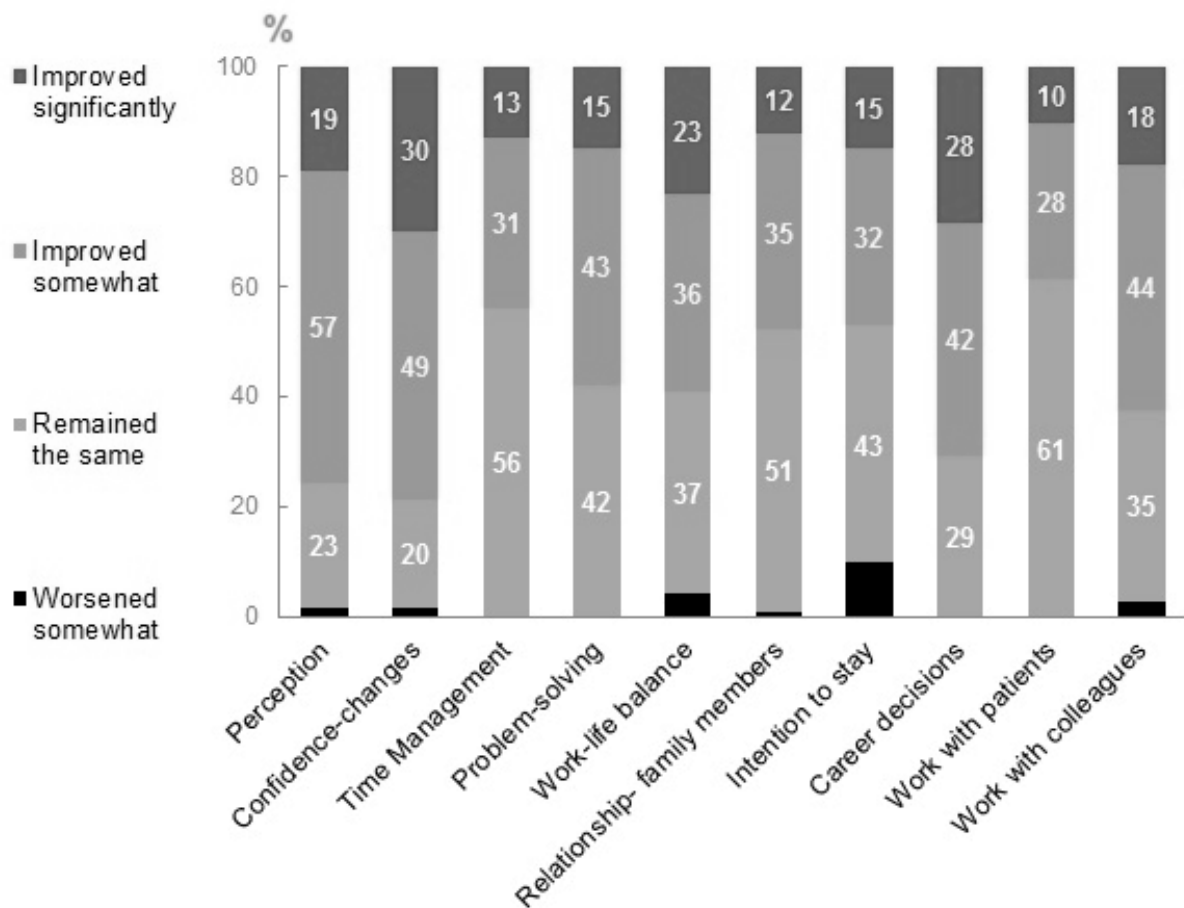

Figure 1   Changes perceived in working life as the result of coaching

Figure 1 represents the results of the analysis of the participants' responses to Q.5: "How the following aspects of your work have changed since starting coaching". This figure illustrates how, according to clients themselves, certain aspects of their working life were changing or remained the same after the period when they undertook coaching. Different shades of grey represent the degree to which the clients perceived the changes. It is clear that the category "worsened significantly" is not present in the figure. The result suggests that the majority of these aspects improved or improved significantly. Only a very small number of responses (21) indicated that some particular aspects of their working life "worsened somewhat". It is interesting to notice that ten of these responses are related to the 'Intention to stay in the current position', which could be interpreted as a positive outcome in some situations when a radical action is beneficial for both the employee and the employer.

Particularly positive perceptions of changes were demonstrated in relation to Perception of values of the client contribution to their current role, Confidence to make changes in the workplace and Ability to make career decisions. In comparison to other factors it appears that the coaching programme was particularly successful in empowering the clients and improving their perception of themselves at work. This indication of changes corresponds with the data of self-efficacy and self-compassion in the Time 1 and Time 2 questionnaires.

Although the results of changes demonstrated in Table 2 and Fig 3 show significant changes in clients who undertook the coaching programme it could be argued that these changes might indicate the influence of a combination of factors other than coaching. Because we had no control group we do not know to what degree the changes that we have seen are due to the coaching that they received. In order to compensate for this issue we included question 6 in our SWRQ, which asked this question directly: "Please indicate the extent to which coaching/mentoring contributed to this change". In relation to each change

from Question 5 the clients could indicate if the change could be attributed to coaching and the degree they believed the coaching influenced this change.

The results of analysis of responses to this question are demonstrated in Fig. 2.
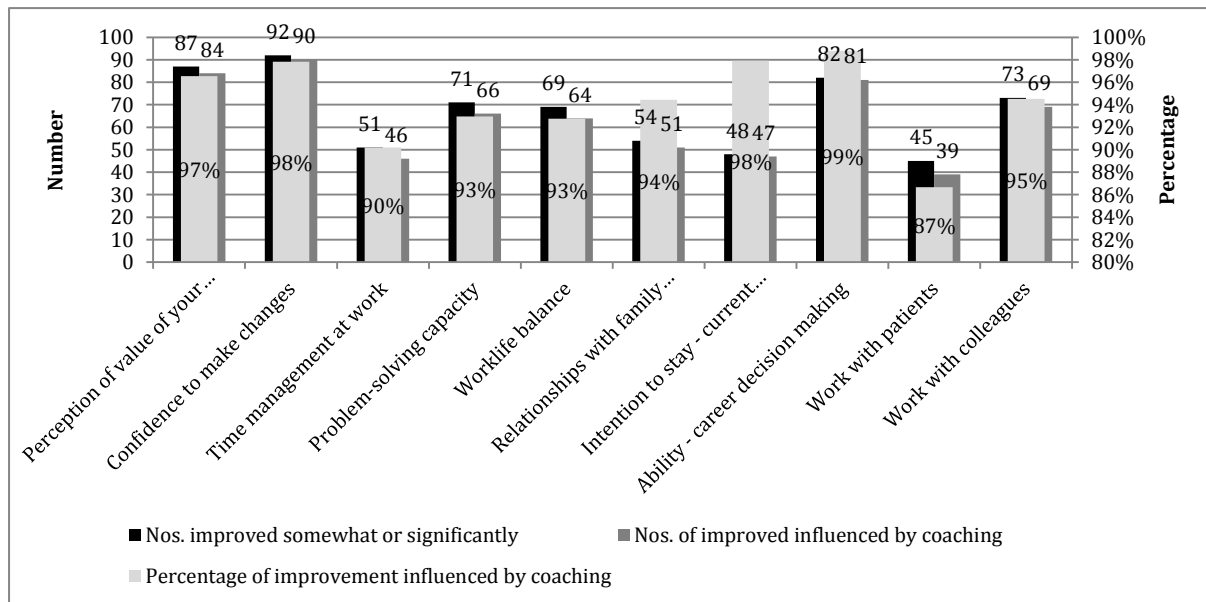

Figure 2   Perception of participants on how changes occurred are attributable to coaching

The black bar in the Fig 2 highlights the number of respondents who felt that there had been some improvement (either somewhat or significant) in the areas highlighted in SWRQ. The dark grey bar highlights the numbers of respondents who felt that the coaching had influenced the positive change that is highlighted by the black bar. The light grey bar sitting in front of two other bars highlights what percentage of positive improvement was attributed to the coaching. For example as was already shown in Fig 1 the three areas of work where respondents felt there had been the most improvement were "Perception of value of your contribution", "Confidence to make changes at work" and "Ability to make career decisions". In these areas of work where there had been improvements, respondents felt that a large percentage of the improvement was due to the coaching that they had received – on average 98% for these three most changed areas of life.

The lowest change is indicated in the area of work with patients, which may indicate that Time 2 responses reflect increased confidence and the way participants felt about themselves, but it may take time to recognize such changes in action particularly with their work with patients. It is also possible that the junior doctors may be fairly remote from patient satisfaction questionnaires (this tends to happen at a departmental level) and therefore find it difficult to notice the effect of their internal changes on patients.

Findings from the open question in SWRQ

Question 9 in the SWRQ asked participants in a word or phrase to describe what difference the coaching and mentoring programme had made to them. All participants responded to this question. Three researchers, first independently and then together, analyzed these responses and identified the following themes. These themes are described in the order of the number of comments considered as representing each theme (from highest to lowest).

Theme (No of comments)	Example of the comments
Confidence (32)	"Substantially increased my confidence in the workplace in the context of being a new consultant joining a well-established senior team".
Change/problem solving (22)	"I can now confidently formulate strategies to help me achieve my goals".
Self-awareness (17)	"…gave me insight into the tools I possess myself to change my work and personal life".
Reflection (16)	"…taught me how to analyse my experiences objectively – reflecting, thinking about things a lot deeper than I usually would"
Work-life balance (12)	"It has improved my perspective on what I am able to achieve at work and so improved my work-life balance significantly. I feel better able to cope as a result."
Seeing things in perspective (8)	"…helped me to see my position, behaviour and current options in better perspective"
Career development (7)	"…focused my ideas of where I want to be in the future and how to influence and use the resources open to me now to reach

	these roles".
Being listened to/sharing (6)	"I was able to safely discuss a very difficult situation at work".
General positive (14)	"An immense difference – turned my life around."
General negative (7)	"…not all problems have a solution".

Table 4. Summary of qualitative analysis

On the whole the qualitative analysis suggests an overwhelmingly positive impact of coaching on clients and a wide variety of the benefits associated with participation in this programme. The overarching patterns of the benefits are:

- Confidence improvement and increased self-awareness

- Specific areas of working life where there was a significant difference as the result of coaching such as career development and work-life balance

- Acquiring a range of skills that could make participants more capable of addressing potential issues, such as the skills of problem-solving, reflection and seeing things in perspective.

It would be unusual if the effect of coaching were universally positive. A small percentage of general negative comments 4.96 illustrate that there are circumstances in which this particular type of intervention is not the best solution. There could be of course other explanations (e.g. not the best match between coach and client); however without further investigation these are only speculations.

Discussion and conclusions

The evaluation described in the report with the support of both quantitative and qualitative methods indicate that the London Deanery coaching programme provides an effective service for their clients. Well-validated measures that were selected for this evaluation project conclusively show that employee engagement, self-efficacy and self-compassion of

the participants significantly improved. It could be argued that the measures selected for this project have been sufficient for the purposes of the evaluation. They appropriately reflect the nature of this programme, which is by definition individually focused. However, as the programme is delivered and paid for by public funding the benefits for the individuals have to be meaningful in the context of the added value to the ultimate users: in this context, patients. Therefore, an additional questionnaire SWRQ was developed with a focus on evaluating changes in the context specific to London Deanery coaching clients. The questionnaire allowed unique information to be elicited about the nature of changes that clients identified as the result of the programme related to their place of work and consequently showed improvements particularly to the aspect of self (confidence in their ability). We believe that this questionnaire would be useful for future evaluations and the data collected in this project can be used for developing the scale for further investigations.

In terms of the design of the evaluation needless to say that projects with features of the randomized control study would be easier to defend in the traditional scientific community. However, there are many reasons to expand this position and we propose to consider at least three of them. The first is simply pragmatic. Considering many issues that constrain the use of RCTs in coaching research (Cavanagh & Grant, 2006; Greif, 2009; Passmore & Fillery-Travis, 2011), not the least the cost-effectiveness of the evaluation itself, we need to develop research designs that are possible to execute. A more positively formulated reason for creating new measures for evaluating coaching programmes relates to giving more status to the participants according to the nature of coaching itself. Coaching is about valuing the voice of the individual, empowering the person and trusting in his/her ability to make a judgment about the issues that reflect his/her life. Research that actively includes the voice of participants in judging the changes they have experienced is better aligned with coaching than research that treats participants as only capable of answering very simple questions leaving the analysis to an 'objective' researcher.

A third and probably the most important reason for being more daring and creative in searching for new approaches to evaluation is based on the acknowledgement that the

outcomes we aim to assess result from the complex interaction of multiple elements in systems with emerging properties. Jones and Corner (2012) recently argued that mentoring should be seen as a case of complex adaptive system (CAS) and their arguments are more than relevant also for coaching. Seeing coaching engagement through this lens brings to the fore many issues that were very difficult to account for in this research. For example, the changes that occurred within the clients are directly and complexly related to their context and it is impossible to isolate all the influences without losing the essence of the process and the layers of meaning for each individual involved. This means that we need to look at research methods that include new ideas different from traditional science because it implies a different view of the world.

In particular we suggest wider use of qualitative methods that can enrich understanding of the effect of coaching. Some of these methods can include, for example, vignettes constructed around actual experiences, by using situations provided by participants before and after coaching. These rich descriptions of actual experiences can be seen as snapshots of the complexity reflected in one particular case without losing the link with its context. We can use the SWRQ for evaluation of coaching programmes as an addition to other methods and an alternative solution when RCT is not possible/appropriate as it includes a self-estimation by the respondents about the extent to which they can attribute each change to the received coaching also without losing the view of the systemic nature of their situations.

We also wish to advocate further building of the theoretical base of coaching by extending evaluation research from the question of effectiveness of coaching to exploring further questions such as the following:

- What elements of coaching in this programme have particularly contributed to positive outcomes?

- Are the changes identified by the clients sustained over a longer period?

- To what type of clients is this programme most suited?

- What difference does the matching process make?

- In what way can the changes identified by the client actually affect their work with patients?

Research into the effect of the coaching programme on a large scale was on the agenda for the London Deanery at the start of this project. Although this could be an important and ambitious undertaking we believe that this current project has provided a positive answer to the question about the effectiveness of the coaching programme on a reasonable scale. In light of these findings we would argue that the questions that aim at improvement of the service which are of a more precise nature, similar to the questions listed above, would be no less important and probably more pertinent for the ultimate stakeholders of the London Deanery's coaching programme.

References

Bandura, A. (1997). *Self-Efficacy: The exercise of control.* New York: Freeman.

Baumeister, R.F. (Ed.) (1999). *The self in social psychology.* Philadelphia, PA: Psychology Press.

Briner, R. (2012) 'Does coaching work and does anyone really care?' *OP Matters,* 16, 4-12.

Cronin, M. & Klimoski, R. (2011). 'Broadening the view of what constitutes "evidence"'. *Industrial and Organizational Psychology*, 4: 1, 57-61.

de Haan, E. and Duckworth, A. (2012). 'Signalling a new trend in executing coaching outcome research'. *International Coaching Psychology Review,* 8:1, 6-19.

Demerouti, E., Bakker, A.B., Janssen, P.P.M. & Schaufeli, W.B. (2001). 'Burnout and engagement at work as a function of demands and control'. *Scandinavian Journal of Work, Environment & Health, 27*, 279-286.

De Meuse, K, Dai, G. & Lee, R. (2009). 'Evaluating the effectiveness of executive coaching: beyond ROI?' *Coaching: An International Journal of Theory, Research and Practice,* 2: 2, 117-134.

Drake, D. (2009). 'Evidence is a verb: A relational view of knowledge and mastery in coaching.' *International Journal of Evidence Based Coaching and Mentoring,* 7: 1, 1-2.

Easton, S. & Van Laar, D. (2013). 'Evaluation of outcomes and Quality of Working Life in the coaching setting'. *The Coaching Psychologist*, *9: 2,* 71-77.

Ellam-Dyson, V. (2012). 'Coaching psychology research: Building the evidence, developing awareness' *OP Matters,* 16, 13-16.

Ely, K., Boyce, L., Nelson, J., Zaccaro, S., Hernez-Broome, G. and Whyman, W. (2010). 'Evaluating leadership coaching: A review and integrated framework'. *The Leadership Quarterly*, 21, 585-599.

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

Franklin, J. & Doran, J. (2009). 'Does all coaching enhance objective performance independently evaluated by blind assessors? The importance of the coaching model and content'. *International Coaching Psychology Review,* 4, 128-144.

Grant, A. (2013). The Efficacy of Coaching. In J. Passmore, D. Peterson & T. Freire (Eds) *The Wiley-Blackwell Handbook of The Psychology of Coaching and Mentoring.* Chichester: John Wiley & Sons.

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

Grief, S. (2007). 'Advances in research on coaching outcomes'. *International Coaching Psychology Review*. 2: 3, 222-245.

Hicks, C. (1998). 'The Randomised Controlled Trial: A Critique.' *Nurse Researcher*, 6: 1, 19-32.

Howard, K, et al (1996). 'Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress'. *American Psychologist*, 51, 1059-1064.

Kirkpatrick, D. (1977). 'Evaluating training programs: Evidence vs. proof'. *Training and Development Journal,* 31: 11, 9-12.

Kühnel, J., Sonnentag, S., & Westman, M. (2009). 'Does work engagement increase after a short respite? The role of job involvement as a double-edged sword'. *Journal of Occupational and Organizational Psychology,* 82, 575-594.

Lane, D. A., & Corrie, S. (2006). *The modern scientist-practitioner: A guide to practice in psychology*. London, UK: Routledge.

Lawrence, P. & Whyte, A. (2014). 'Return on investment in executive coaching: a practical model for measuring ROI in organisations'. *Coaching: An International Journal of Theory, Research and Practice,* 7: 1, 4-17.

Luszczynska, A., Scolz, U. & Schwarzer, R. (2005). 'The General Self-Efficacy Scale: Multicultural Validation Studies'. *The Journal of Psychology: Interdisciplinary and Applied.* 139: 5, 439-457.

Neff, K. D. (2009). 'The role of self-compassion in development: A healthier way to relate to oneself'. *Human Development,* 52*,* 211-214*.*

Neff, K. D. & Lamb, L. M. (2009). 'Self-Compassion'. In S. Lopez (Ed.), *The Encyclopedia of Positive Psychology,* Blackwell Publishing.

Neff, K. D. & Vonk, R. (2009). 'Self-compassion versus global self-esteem: Two different ways of relating to oneself'. *Journal of Personality,* 77, 23-50*.*

Neff, K. D. (2003). 'Development and validation of a scale to measure self-compassion'. *Self and Identity,* 2, 223-250.

Orlinsky, D., Grawe, K. & Parks, B. (1994). 'Process and outcome in psychotherapy', In A. Bergin & S. Garfield (eds), *Handbook of psychotherapy and behavior change,* (4 ed) New York: John Wiley.

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

Passmore, J. (2008). (Ed.) *Psychometrics in Coaching,* London: Kegan Page.

Passmore, J. & Joao Velez, M. (2014). Training Evaluation, in K. Kraiger, J. Passmore, A. Dos Santos, & S. Malvezzi (eds) *The Wiley and Blackwell Handbook of the Psychology of Training, Development and Performance Improvement,* Chichester: Wiley and Blackwell, pp. 136-153.

Salanova, M., Agut, S. & Peiró, J.M. (2005). 'Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate'. *Journal of Applied Psychology,* 90, 1217-1227.

Schaufeli, W.B. & Bakker, A.B. (2004). 'Job demands, job resources and their relationship with burnout and engagement: A multi-sample study'. *Journal of Organizational Behavior, 25*, 293- 315.

Schaufeli, W.B., Salanova, M., Gonzalez-Romá. V. & Bakker, A.B. (2002). 'The measurement of engagement and burnout: A two-sample confirmatory factor analytic approach'. *Journal of Happiness Studies,* 3, 71-92.

Scherbaum, C., Cohen-Charash & Kern, M. (2006). 'Measuring General Self-Efficacy: A comparison of Three Measures Using Item Response Theory'. *Educational and Psychological Measurement*. 66: 6, 1047- 1061.

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,* Windsor, UK: NFER-NELSO.

Stajkovic, A. & Luthans. F. (1998). 'Self-Efficacy and Work-Related Performance: A Meta-Analysis'. *Psychological Bulletin,* 124: 2, 240-261.

Strauss, A. and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.

Tashakkori, A. and Teddli, C. (Eds), (2010). *Sage Handbook of Mixed Methods in Social & Behavioural Research*. London: Sage.

Theeboom, T., Beersma, B. and van Vianen, A. (2013). 'Does coaching work? A meta-analysis on the effects of coaching on individual level outcomes in an organisational context'. *The Journal of Positive Psychology*, 9: 1, 1-18

Williams, B.A. (2010). 'Perils of evidence-based medicine'. *Perspectives on Biology and Medicine, 53*: 1, 106–120.