The unsolved value of executive coaching: A meta-analysis of outcomes using randomised control trial studies

Daniel Burt, School of Psychology and Exercise Science, Murdoch University, Australia

Zenobia Talati, School of Psychology and Speech Pathology, Curtin University, Bentley, Australia

Email: daniel.joshua.burt@gmail.com

Abstract

Methodology and research supporting coaching's effectiveness has not kept up with its growth and demand. The current literature on coaching is lacking sufficient empirical rigour and does not meet the standard required for mixed methods design. This meta-analysis investigated the outcomes of coaching, and potential moderating effects of other factors, using only randomised control trial studies. Outcomes studied included performance, well-being, coping, work attitudes, and goal-directed self-regulation. There were no moderating effects identified from participant age, type of measure, or author(s). The results showed that overall coaching has a moderate significant positive effect on coachees, $\hat{p} = 0.42$, which indicated that coaching is effective for individuals.

Key Words: executive coaching, randomised control trial, meta-analysis, outcomes, development.

Introduction

Executive coaching has developed from a scarcely used leadership development practice in the late 1980s to its current position among the most effective and commonly used tools benefiting leaders (Coutu & Kauffman, 2009; McGovern et al., 2001; Tobias, 1996). Given that the principal concern for those practising coaching is the presence of untrained coaches (International Coach Federation, 2012), the influence of psychologists and their accompanying understanding of best practice methodology and assessment may help improve the quality of coaching provided by all coaches. Beyond the significant financial investment made by participating companies, coaching takes an individual away from their work with the premise that there will be a return on the investment and consequently there is a duty to prove that it actually works. Despite its popularity, the question remains: What makes it effective?

Psychological Theories Relevant to Coaching

There are many psychological theories underpinning the coaching process, such as goal setting, feedback, and identification of more severe underlying issues in clients. For example, a meta-analysis (Kluger & DeNisi, 1996) found that, on average, feedback interventions improved performance. However, certain types of tasks and motives changed this effect. Therefore, leveraging this knowledge of which types of feedback are relevant for certain situations could be instrumental in improving the client's outcomes.

Another area where psychology can add value to coaching is through empirical rigour. Within the coaching literature, there is a distinct lack of best practice approaches when it comes to assessing the efficacy of the coaching intervention (De Meuse et al., 2009). Recent reviews from the USA and UK suggest that only a small proportion of organisations are accurately evaluating the outcomes of coaching; less than 10 per cent in USA and only 19 per cent in the UK (Bolch, 2001; Hay Group, 2002). If clients and organisations merely observe improvements at face value without objective measurement throughout the coaching process, there cannot be any proof of that development.

While there has been a marked increase in the number of qualitative reviews and peer-reviewed journal publications on coaching (Grant & Cavanagh, 2004), little progress has been achieved by way of pre- post studies to demonstrate improvements following a coaching intervention. There is also a distinct lack of control groups in pre- post executive coaching assessments. De Meuse, Dai, and Lee (2009) claim that including control groups is not possible due to the very nature of coaching objectives, as they constantly change and adapt to changing requirements. However, others such as Stober and Grant (2006) have attempted, with success, to randomly include control participants in coaching research. According to Grant, Passmore, Cavanagh, and Parker (2010), since the first known randomised control study conducted by Deviney (1994) there had been eleven between-subjects studies that had successfully compared control to experimental groups in research (see Grant, Green, & Rynsardt, 2013; Moen & Federici, 2012).

Regardless of whether a worker has received coaching or not, learning and behavioural transformation are likely to occur, albeit at theoretically slower rates than in the presence of coaching. As many biases exist within a one to one relationship it is critical to measure interventions against randomised control groups. Research that does not include a comparison group is at risk of applying too much credit to coaching for the measured changes. For example, Smither, London, Flautt, Vargas and Kucine (2003) found that measuring the effect of coaching against controls returned positive yet small improvements. At the time, this was the only published research to measure outcomes for both the experimental (received coaching) and control (did not receive coaching) groups.

De Meuse et al (2009) conducted the first known meta-analysis on executive coaching outcomes that included randomised control trial studies, however the authors were only able to acquire six suitable papers and subsequently included other, non-randomised control trial studies in their meta-analysis. Despite the paucity of the studies, the authors found that overall coaching had a positive effect, which was strengthened when tied to specific objectives. Later Theeboom et al (2014) produced a meta-analysis that included eight control studies focusing on quantitative data for outcomes; however these papers included post-test only measures (e.g., Poepsel, 2011). Theeboom et al (2014) established that studies utilising exclusively within-subjects designs reported considerably stronger effect sizes than the studies that included control groups, which suggested that research using a within-subjects methodology may be over estimating the effect of coaching. Theeboom et al (2014) also noted the shortage of rigorous studies available.

The Present Study

The purpose of the present thesis is to critically evaluate and review the existing research on the effectiveness of coaching through a meta-analysis of rigorous randomised control trials. Previous systematic analyses on coaching have called for larger samples of rigorous research (De Meuse et al., 2009) or further investigation into what makes coaching effective (Theeboom et al., 2014). The following questions will be addressed:

Q1: Which outcomes are most affected by coaching?

Q2: What factors influence (or moderate) the effect of coaching?

It has been suggested that certain age groups face specific challenges in life that may change the coaching process (Green et al., 2007), therefore age will also be examined as it may potentially moderate the effectiveness of coaching. This study will add to the existing literature and previous meta-analytical reviews (De Meuse et al., 2009; Theeboom et al., 2014) by:

1) including only randomised control studies with pre-test and post-test data for control and experimental groups;
2) taking the methodological quality of the primary studies into account;
3) including recent studies published after the most recent meta-analysis (2012 – 2014);
4) including unpublished dissertations that meet the selection criteria; and
5) applying clear inclusion criteria for the type of interventions and study design.

Method

A literature search was carried out using PsychINFO, Google Scholar, and Proquest databases to identify relevant studies published up to June 30, 2014. The following terms were used in the search: "executive coaching randomised control", with additional terms "thesis" or "dissertation" used to refine the search. In addition to the term "executive", similar coaching related labels "leadership", "developmental", "business", and "life" were included. The terms "outcome" and "effect" were used to target quantitative studies. This search strategy yielded a total of 1,055 articles. The studies cited in previous reviews and meta-analyses (i.e., De Meuse et al., 2009; Grant, Passmore et al, 2010; Greif, 2007; Theeboom et al., 2014) were cross-checked and included if relevant.

Leading coaching academics were contacted via email in search of soon to be released publications that could be included in this study. This search strategy did not yield any relevant articles. Details on how many studies were removed at each stage can be found in Table 1.

		Search	
Phase	Summary	Theses	Journal articles
1	Initial keyword search	1,055	
2	Potential theses and published studies identified via electronic database combing, citation searching, and email requests	40	103
3	Articles remaining after selection criteria applied: - Randomised control trial - Pre-test, post-test means and standard deviations (SD) provided - One to one coaching as the independent variable - Written in English	5	7
4	Articles remaining after duplicate studies removed	5	6
5	Total remaining studies	11	

Table 1. Search strategy for the inclusion of theses and published studies in the meta-analysis.

Inclusion and Exclusion Criteria

Studies eligible for inclusion were screened in two phases. To pass the first stage of screening, studies were required to have a title and abstract which suggested that pre-test, post-test, and control measures were taken. In the second stage, studies were assessed based on a review of the complete article. The inclusion criteria were as follows:

- sufficient statistics were reported in the article to enable calculation of standardised effect sizes, such as pre-test and post-test means, standard deviation, group sample size, and alpha coefficients;
- the study needed to be a randomised control trial;
- at least one self-reported or objective outcome measure was reported, such as well-being, depression, goal attainment, or self-rated job performance;
- a coaching method design equivalent to Grant's (2003) definition was used: *"…a collaborative solution focused, result-orientated and systematic process in which the coach facilitates the enhancement of life experience and goal attainment in the personal and/or professional life of normal, nonclinical clients"* (p. 254);
- the coaching intervention was at least partially completed in a one-to-one setting, with an external coach.

To avoid duplication of data, theses that were eventually published were included as the published article. However, if the thesis had additional results, the data from the thesis were included instead (e.g., Green, 2004; Spence, 2006).

Calculating Effect Size

In a meta-analysis the effect sizes from measured outcomes in different studies are converted into a standardised effect size that is no longer positioned on the original measurement scale and can therefore be compared with outcomes from other scales.

The present meta-analysis calculated effect sizes using a method recommended by Hoffman and Smits (2008). In order to standardise the data across each randomised control trial, the mean difference between the pre-test and post-test measures was calculated for the experimental and the control groups. This was used to calculate the Hedges' *g* effect size and its 95% confidence interval. This effect size is a variation on Cohen's *d* that corrects for biases due to small sample sizes (Hedges & Olkin, 1985) and is calculated using the following formula:

$$g = \frac{\bar{\Delta}_{\text{experimental}} - \bar{\Delta}_{\text{control}}}{\sqrt{\dfrac{(\text{n}_{\text{experimental}}-1)SD^2_{\text{experimental}} + (\text{n}_{\text{control}}-1)SD^2_{\text{control}}}{(\text{n}_{\text{total}}-2)}}} \text{ X } \left(1 - \frac{3}{4(\text{n}_{\text{experimental}} + \text{n}_{\text{control}})-9}\right)$$

In the equation above, $\bar{\Delta}$ is the mean pre- to post-test change, SD is the standard deviation of post-test scores, n is the sample size, experimental refers to the experimental condition, and control refers to the control condition.

According to Hedges (2008), the standardised mean difference and related effect sizes such as those found in the studies included in this meta-analysis are normally examined by calculating the change in means on a particular dependent variable either from pre- to post-test or between the experimental and control group (both at post-test). As a preference, generalised eta squared would have been calculated as it can be used for more than two sets of observations to compare changes both within-groups and between-groups (Lakens, 2013). However the data required to calculate this was not provided in most studies. Previous meta-analyses have focused on post-test only comparisons of the experimental and control group (Theeboom et al., 2014). Our approach was deemed more representative of true effect size as it took into account the between-subjects (i.e., experimental versus control), and within-subjects (i.e., pre-test versus post-test) differences. Importantly it takes into account how people *change* as a result of coaching rather than just looking at where they end up after the coaching.

Whether an effect size is considered small or large is open to interpretation. According to Cohen (1988), descriptions include approximations around small (*d* = 0.2), medium (*d* = 0.5), and large (*d* = 0.8). However in the context of a random effects model for meta-analysis, estimations of confidence limits for the overall mean effect are used to explain the effect size (Schmidt & Hunter, 1995).

Meta-analytic Procedure and Statistical Analyses

MIX 2.0 - Professional software for meta-analysis in Excel was used to perform the meta-analysis. This application was chosen from a systematic comparison of software dedicated to meta-analysis, where MIX scored highest on the overall usability when compared to four other prominent software packages, as well as featuring the relevant analytical feature comparison of weighting according to alpha coefficients, confidence intervals and small study effect corrections (Bax, Yu, Ikeda, & Moons, 2007).

Similar to Theeboom et al (2014), the overall weighted effect size for the meta-analysis was calculated using the Hedges and Olkin (1985) method. Meta-analysis specialists generally regard the Hedges and Olkin random effects method as the most conservative approach as it corrects for sample size, measurement error and range restriction by removing the ability to inflate effect size estimates, manipulate statistical corrections for artifactual variance sources such as measurement error and restriction of range (Borenstein, Hedges, Higgins, & Rothstein, 2011). As recommended by Borenstein et al (2011) the more conservative random-effect model was adopted for the meta-analysis. Compared to the fixed effect model, the random-effect approach permits that the true effect size differs from study to study based on both the variability of the independent variable, such as intensity or duration of intervention, and differences in the samples of the research population such as gender, age, and coaching experience (Borenstein et al., 2011; Hedges, 1994). This approach is generally considered best practice and recommended by Field (2001). The overall weighted mean effect size for all included studies will be represented by $\hat{p}$, while g will represent the weighted effect sizes of individual studies. From the studies included in this paper there were some cases where reliability values were not reported for the dependent variables. For these cases the alpha coefficients reported in the initial testing of these instruments were utilised. Where a range was provided the median was taken, e.g., if alpha coefficients ranged from 0.80-0.92, an alpha of 0.86 was employed.

Moderator Analyses

As recommended by Higgins and Thompson (2002), heterogeneity between studies was measured by calculating both the classical Q statistic and the I^2 statistic. The most commonly used homogeneity statistic is Cochran's Q (Cochran, 1954), which calculates a weighted sum of the square distances of the observed effects from the null hypothesis of equality of the effects. The Q statistic provided a significance test for between-study heterogeneity, whereas the value for I^2 represented the percentage of between-study variance in effect sizes that can be attributed to between-study heterogeneity rather than within-study variability (Borenstein et al., 2011). Moderating variables were investigated using subgroup analyses for sets of studies that differed in terms of the study, author(s), age of participants, published and unpublished data, and outcome variables. Moderating variables are considered to be present when the value for I^2 is above or close to 50 percent (Higgins & Thompson, 2002). The included studies and measures are presented in Table 2.

Results

Study Selection

The screening process resulted in a total of 11 published and unpublished studies that were included in the final analysis. Green (2004) contained data that was later published in Green et al., (2006) and Spence (2006) contained data that was later published in Spence and Grant (2007).

Overall Effect Size and Homogeneity

Article/Thesis Author(s)	Thesis/ Published	Country of Origin	Outcome Measure	Instrument(s)	*n*	Participant Gender	Average duration (in weeks)	Average number of sessions
Richardson (2010)	Thesis	USA	Goal Attainment Satisfaction with Life Working Alliance	GAS SWLS WAI	18	Data not included	6	Data not included
Finn (2007)	Thesis	AUS	Self-Efficacy Developmental Support Positive Affect Openness to New Behaviours Developmental Planning	TSLES Development Scale PAS Openness to new behaviours Developmental Planning	23	21.74% female	12	6
Green (2004)/ Green, Oades, & Grant (2006)^	Thesis/ Publication^	AUS	Satisfaction with Life Hope Positive and Negative Affect Depression, Anxiety, Stress Well-Being Personal Striving Goal Striving Progression	SWLS DASS PANAS HTS PS GSP SPWB	56	75% female	10	Data not included
Grant (2001)	Thesis	AUS	Depression, Anxiety, Stress Study Process Self-Control Motivated Strategies for Learning Test Attitude Private Self-consciousness	DASS SPQ SCS MSLQ TestAnx PrSCS	20 18 24	60% female 72% female 50% female	Data not included	6
Spence (2006) / Spence, & Grant (2007) ^	Thesis/ Publication^	AUS	Positive and Negative Affect Satisfaction with Life Well-Being	PANAS SWLS SPWB	63	74.6% female	Data not included	10

Table 2. Characteristics of included studies and measures (continued over page).

Article/Thesis Author(s)	Thesis/ Published	Country of Origin	Outcome Measure	Instrument(s)	*n*	Participant Gender	Average duration (in weeks)	Average number of sessions
Bozer & Sarros (2012)	Publication	AUS	Self-reported Job Performance Self-awareness Job Affective Commitment Career Satisfaction Supervisory-related task performance Supervisor-report Job Performance	JPS SIS JACS CSS SRTPS	197	47.5% female	11	6.8
Grant, Curtayne, & Burton (2009)	Publication	AUS	Goal Attainment Cognitive Hardiness Depression, Anxiety, Stress Well-being	GAS CHS DASS WWBI	41	92.68% female	9	10
Grant, Green, & Rynsaardt (2010)	Publication	AUS	Goal Attainment Cognitive Hardiness Depression, Anxiety, Stress Well-being	GAS CHS DASS WWBI	44	70.45% female	20	10
Green, Grant, & Rynsaardt (2007)	Publication	AUS	Hope Trait Cognitive Hardiness Depression, Anxiety, Stress	HTS CHS DASS	56	100% female	10	10
O'Connor & Cavanagh (2012)	Publication	AUS	Goal Attainment Well-Being	GAS SPWB	102	55% female (including coaches)	18	8

Table 2. Characteristics of included studies and measures.

Note. CHS = Cognitive Hardiness Scale; CSS = Career Satisfaction Scale; DASS = Depression, Anxiety and Stress Scale; GAS = Goal Attainment Scaling; GSP = Goal Striving Progression; HTS = Hope Trait Scale; JACS = Job Affective Commitment Scale; JPS = Job Performance Scale; MSLQ = Motivated Strategies for Learning; PANAS = Positive and Negative Affect Scale; PAS = Positive Affect Scale; PrSCS = Private Self-Consciousness Scale; PS = Personal Striving; SCS = Self-control Schedule; SIS = Self Insight Scale; SPQ = Study Process Questionnaire; SPWB = Scales of Psychological Well-being; SRTPS = Supervisory-related task performance; SWLS = Satisfaction with Life Scale; TestAnx = Test Attitudes Questionnaire; TSLES = Transformational Leadership Self-Efficacy Scale; WAI = Working Alliance Inventory ; WWBI = Workplace Well-being Inventory.

The random effects meta-analysis of executive coaching against control groups yielded an overall weighted effect size for all outcomes that was in the small to medium range, reflecting an advantage of coaching over control groups, $\hat{p}$ = 0.42, 95% CI, 0.35 – 0.50, p <0.001, as presented in Table 3. The homogeneity in effect sizes was not statistically significant although moderate in scale, Q = 126.95; p = 0.081; I^2 = 16.50. According to Higgins and Thompson (2002), variance of this magnitude does not warrant examination of moderator variables. However, given the exploratory nature of this study, further investigation of moderators took place.

			CI (95%)				
k	n	$\hat{p}$	Lower	Upper	p-value	Q	I^2
11	696	0.42	0.35	0.50	<0.001	126.95	16.50%

Table 3. Overall effect size and homogeneity.

Note. k = number of studies included in the analysis; n = total sample size in k studies; $\hat{p}$ = overall weighted mean effect size; CI = 95% random effects confidence intervals; Q = Cochran Q statistic; I^2 = the percentage of between-study variance in effect sizes that can be attributed to between-study heterogeneity.

Within-Subjects Mean Difference

Table 4 contains the within-subjects effect sizes for the experimental and control groups. Overall, there was a significant positive change across the experimental groups, $\hat{p}$ = 0.40, 95% CI, 0.32 – 0.48, p < 0.001, whereas the control groups did not change beyond reasonable doubt, $\hat{p}$ = 0.04, 95% CI, -0.03 – 0.10, p = 0.281.

The confidence intervals for the experimental group did not overlap with control group. According to Higgins and Thompson (2002) this result indicates that the effect was greater within the experimental group.

				CI (95%)		
Group	k	n	g	Lower	Upper	p-value
Experimental	11	284	0.40	0.32	0.48	<0.001
Control	11	412	0.04	-0.03	0.10	0.281

Table 4. Effect sizes for the within-groups measures.

Note. n = total sample size; g = Hedges' g; CI = 95% random effects confidence intervals.

Between-study Effect Sizes and Homogeneity

Table 5 contains the weighted effect sizes (aggregated over outcomes) per study. Four research papers returned confidence intervals that included zero (see Bozer & Sarros, 2012; O'Connor & Cavanagh, 2012; Richardson, 2010; Spence & Grant, 2007), which indicated that the effect was weak and could potentially be 0.00 (Schmidt & Hunter, 1995).

The Effect of Coachee Age on the Outcome Variables

As a range of ages were noted throughout the studies, the average age from each study was investigated as a potential moderator. Participants in studies with an average age younger than 30 years, g = 0.46, 95% CI, 0.36 – 0.57, $p < 0.001$, were slightly more receptive to coaching than those studies with an average age of 30 years and older, g = 0.37, 95% CI, 0.26 – 0.48, $p < 0.001$. Both groups responded positively to the intervention, and a t-test revealed that there was no significant difference between the two groups, $t(100) = 0.31$, $p = 0.758$. While the age group 30 years and above did yield a significant homogeneity statistic, the proportion of total variation in the estimates of treatment effect that is due to heterogeneity between studies was below the recommended threshold (Higgins & Thompson, 2002), $Q = 60.66$; $p = 0.025$; $I^2 = 32.41\%$.

			CI (95%)				
Study	n	g	Lower	Upper	p-value	Q	I^2
Bozer & Sarros (2012)	197	0.08	-0.31	0.47	0.683	19.82	74.78%
Finn (2007)	23	0.96	0.40	1.53	0.001	5.66	29.36%
Grant (2001)	62	0.48	0.36	0.60	0.000	41.56	0.00%^
Grant, Curtayne, & Burton (2009)	41	0.46	0.18	0.74	0.001	3.47	0.00%^
Grant, Green, & Rynsaardt (2010)	44	0.42	0.15	0.69	0.002	2.30	0.00%^
Green, Oades, & Grant (2006)	56	0.64	0.47	0.81	0.000	8.94	0.00%^
Green (2004)a	56	0.36	0.03	0.68	0.030	0.27	0.00%^
Green, Grant, & Rynsaardt (2007)	56	0.42	0.02	0.82	0.039	12.20	59.02%
O'Connor & Cavanagh (2012)	102	0.30	-0.19	0.79	0.230	0.00	0.00%^
Richardson (2010)	18	0.62	-0.09	1.34	0.087	0.25	0.00%^
Spence & Grant (2007)b	41	0.15	-0.07	0.36	0.187	4.32	0.00%^

Table 5. Weighted effect sizes per study aggregated over outcomes.

Note. n = total sample size; g = Hedges' g; and CI = 95% random effects confidence intervals; Q = Cochran Q statistic; I^2 = the percentage of between-study variance in effect sizes that can be attributed to between-study heterogeneity.

[a] *Green (2004) contained data that was later published in Green, Oades, and Grant (2006);*

[b] *Spence (2006) contained data that was later published in Spence and Grant (2007.).*

^ *The I^2 was truncated to zero because the Q statistic used for the computation of I^2 was smaller than its degrees of freedom.*

Effect Sizes per Outcome Category

Table 6 displays the results for all outcome categories: attitudes, coping, performance, self-regulation, and well-being. Attitudes, g = 0.78, 95% CI, 0.54 - 1.03, $p <$ 0.001, coping, g = 0.68, 95% CI, 0.33-1.03, $p <$ 0.001, self-regulation, g = 0.43, 95% CI, 0.30 - 0.56, $p <$ 0.001, and well-being, g = 0.41, 95% CI, 0.31 - 0.50, $p <$ 0.001, were all found to be positively influenced by coaching. These results suggest that coaching interventions have significant positive effects on all outcome categories. It is important to note that only one study (Bozer & Sarros, 2012) included performance-related ratings which comprised of self-rated and observer-rated measures. While other authors did include performance ratings, these measures were not assessed for internal consistency and therefore could not be included in this analysis. Measures that did not have an alpha coefficient were omitted, for example, Spence and Grant's (2007) use of Goal Attainment Scaling was not included as there was no alpha coefficient reported.

				CI (95%)				
Outcome	*k*	*n*	*g*	Lower	Upper	*p*-value	*Q*	I^2
Attitudes	5	215	0.78	0.54	1.03	0.000	1.93	0.00%^
Coping	4	164	0.68	0.33	1.03	0.000	0.87	0.00%^
Self-regulation	3	282	0.43	0.30	0.56	0.000	43.83	0.00%^
Well-being	11	696	0.41	0.31	0.50	0.000	45.72	1.58%

Table 6. Weighted effect size of coaching interventions on all outcome categories.

Note. k = number of studies included in the analysis; n = total sample size in k studies; g = Hedges' g; CI = 95% random effects confidence intervals; and Q = Cochran Q statistic. I^2 = the proportion of total variation in the estimates of treatment effect that is due to heterogeneity between studies. Performance was not included as it only contained one study. ^ = the I^2 was truncated to zero because the Q statistic used for the computation of I^2 was smaller than its degrees of freedom.

Table 7 contains the measures used to determine the outcomes. The results suggest that overall workplace specific well-being improves following the coaching intervention. The greatest improvement was found in Test Attitude Inventory, g = 0.95, 95% CI, 0.42 – 1.49, $p <$ 0.001, while the smallest effect size with confidence intervals overlapping zero was found in the Positive Affect Scale, g = 0.35, 95% CI, -0.24 – 0.94, p = 0.244. These results will be further reflected upon in the discussion section.

				CI (95%)		
Measure	*k*	*n*	*g*	Lower	Upper	*p*-value
			Coping			
CHS*	3	120	0.66	0.29	1.03	<0.001
			Well-being			
DASS-A*	7	245	0.28	0.03	0.54	0.028
DASS-D*	7	244	0.47	0.21	0.73	<0.001
DASS-S*	7	231	0.40	0.10	0.69	0.009
NAS	2	87	0.45	-0.20	1.10	0.175
PAS	3	104	0.35	-0.24	0.94	0.244
SPWB*	3	189	0.41	0.22	0.60	<0.001
SWLS	3	103	0.36	-0.03	0.76	0.070
WWBI*	2	85	0.44	0.01	0.87	0.046
			Attitudes			
HTS*	2	99	0.76	0.46	1.06	<0.001
TestAnx*	3	62	0.95	0.42	1.49	<0.001
			Self-regulation			
MSLQ*	3	62	0.74	0.22	1.26	0.005
PrSCS*	3	67	0.37	0.20	0.54	<0.001
SCS*	3	62	0.57	0.00	1.14	0.049
SPQ*	3	186	0.41	0.06	0.76	0.022

Table 7. Weighted effect size of measures.

Note. Only measures used in two or more studies were included. *k* = number of studies included in the analysis; *n* = total sample size in *k* studies; *g* = Hedges' *g*; and CI = 95% random effects confidence intervals. CHS = Cognitive Hardiness Scale; DASS = Depression, Anxiety and Stress Scale; HTS = Hope Trait Scale; MSLQ = Motivated Strategies for Learning; NAS = Negative Affect Scale; PAS = Positive Affect Scale; PrSCS = Private Self-Consciousness Scale; SCS = Self-control Schedule; SPQ = Study Process Questionnaire; SPWB = Scales of Psychological Well-being; SWLS = Satisfaction with Life Scale; TestAnx = Test Attitudes Questionnaire; WWBI = Workplace Well-being Inventory. * = statistically significant results overall.

Publication Bias

As illustrated in Figure 1, there were no noticeable signs of publication bias. Furthermore a sub-group analysis of the post-test outcomes revealed that unpublished theses, g = 0.39, 95% CI, 0.28 – 0.50, p < 0.001, returned very similar results to published articles, g = 0.41, 95% CI, 0.28 – 0.53, p < 0.001. However it should be noted that some of the theses were eventually published within peer-reviewed journals.

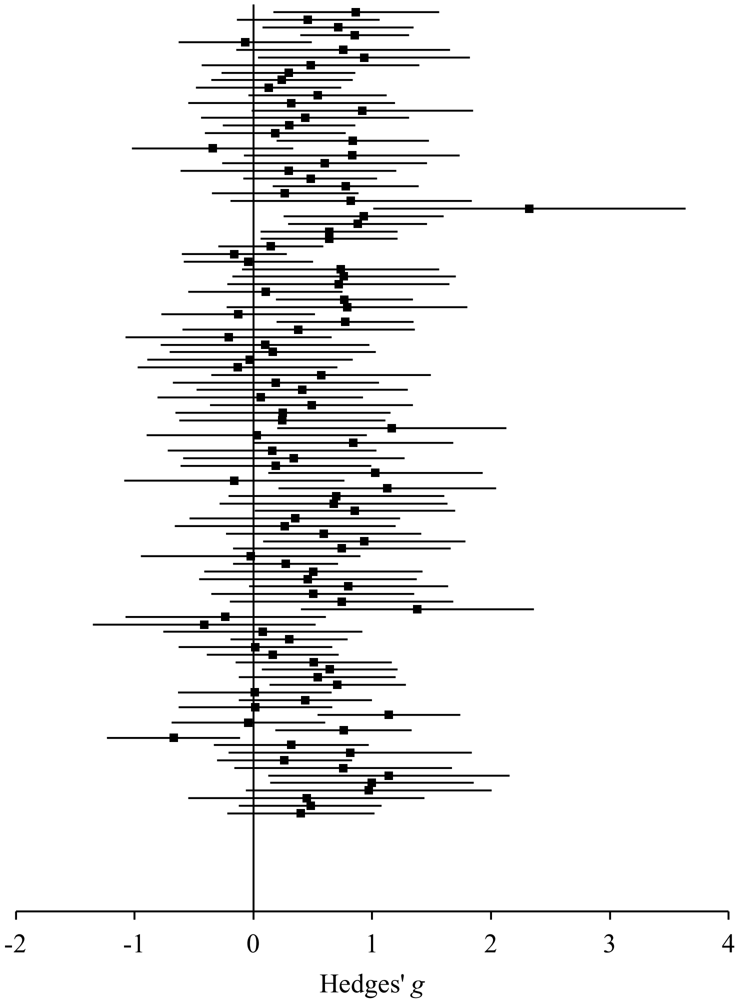

Figure 1. Funnel plot containing Hedges' g with 95% confidence intervals.

Discussion

The current literature on coaching supports its efficacy in improving workplace performance and well-being. Prevailing meta-analyses have focused on using either quasi-experimental field study data or exclusively post-test outcomes to determine overall effect sizes (see De Meuse et al., 2009; Theeboom et al., 2014). The findings from these analyses do not take into account the control group's outcomes, nor the differences between the control group and experimental group's pre-test scores. Limiting the current meta-analysis to include solely randomised control trial (RCT) studies removed some of these methodological issues and provides a more accurate picture of the effects of coaching.

The purpose of the present thesis was to critically evaluate and systematically review the existing research on the effectiveness of coaching and reach a conclusion through the use of meta-analysis. Associations between the coaching intervention and multiple types of outcomes were investigated using best practice methodology (i.e., exclusively RCT studies). The results show that coaching has a moderate positive effect on well-being, work-related attitude, coping strategies, and self-directed goal attainment. In summary, the meta-analysis revealed that coaching is an effective tool for improving individuals' perception about themselves and their workplace. In practice, individuals with improved well-being will provide the organisation they work for with a positive outcome, however coaching assignments are taken on with the intention of developing specific competencies to improve performance, not just well-being.

The overall effect size findings of the current research were considerably more conservative than recent meta-analyses that reported a medium to strong effect size overall (Theeboom et al., 2014) and a very strong effect size (De Meuse et al., 2009). The notable difference in effect size between the current study and previous meta-analyses can be attributed to the difference in methodology. For example, unlike Theeboom et al (2014) who only included post-test data, the current study controlled for pre-test data in the analysis. Including both pre-test and post-test data allows for the assessment of change in measures resulting from the intervention as it accounts for the differences between the control and experimental groups prior to, as well as after, the intervention.

Another factor giving rise to the difference in effect could be sample size. For instance, ever since De Meuse et al (2009) first raised the issue of the scarcity of robust data, the effect of coaching on outcomes as a topic of research has grown in popularity. Such progress is apparent in the increase in RCT studies available, growing from four available papers in 2009 to the eleven included in this study.

The investigation into coachee age revealed that younger participants had slightly stronger outcomes in contrast to their older peers, however not to a level that warranted further investigation.

A between-measures moderator analysis was conducted to further expand on previous research into coaching outcomes. The largest effect was found in the Test Attitudes Questionnaire (Spielberger, 1980), which measures an individual's anxiety in relation to a test. All other measures included within the study only showed a moderate improvement or less. Remarkably, the Positive and Negative Affect Scale (Watson, Clark, & Tellegen, 1988) did not perform nearly as consistently as similar measures like the Depression, Anxiety, and Stress Scale (Lovibond & Lovibond, 1995). Although moderately correlated (Crawford & Henry, 2004), these two scales returned noticeably different results, which demonstrates the importance of choosing the right psychological measure for the situation. These findings reiterate how crucial it is to have trained/accredited psychologists' input when assessing change in coachee behaviour. For this reason it is critical to coaching's pursuit of being considered a profession that at least some psychological contribution be considered when designing coaching solutions.

Limitations

Although many of the methodological problems of meta-analytic studies were avoided, there remained some notable weaknesses to the current study. The first limitation was that one research group related to the University of Sydney (e.g., Grant, Green, & Rynsaardt, 2010) published seven of the eleven theses and articles included herein. Their involvement in over 50 per cent of the studies could perhaps be deemed as having too much influence over the outcomes of the meta-analysis. However the fact that studies from this group were within the outcome range of the others suggest that no bias was present.

Secondly, coaching interventions and outcomes across the included studies varied greatly, which limited the amount of direct comparisons that could be made for a meta-analytic synthesis. Furthermore, employing strict selection criteria resulted in a relatively small number of studies included in the final analysis. Although there was no evidence for publication bias and the inclusion of eleven studies was above the minimum of two for meta-analysis (Sterne, Egger, & Moher, 2008), further research is still required to substantiate these results.

Future Research

The findings of the current study could be used by coaches to form balanced and honest advice to their clients about the accurate and reliable benefits of coaching. More rigorous research designs (featuring randomisation to a control group) are needed to support coaching as an evidenced based profession. As Grant, Passmore et al (2010) discussed in a review of the coaching literature, there is a strong emphasis on descriptive studies that investigate practice-related issues as opposed to proving the effectiveness of the coaching intervention. It is broadly understood that the research design with the highest methodological rigor is an RCT (Campbell & Stanley, 1963). While within-subject studies remain helpful when focusing on the complexities of the coach-coachee relationship there is now strong cause for evidence-based practice to set the standard for coaches. A coach's guidance can only account for a finite amount of coachee improvement. If researchers employ within-subjects research methods, a leader's experience, values, special talents and interests are not adequately accounted for. To consider the effectiveness of coaching in the wider context of the whole person, RCTs represent the best technique to measure the effect of coaching on leadership functioning. Combining the RCT design with taking pre-test, post-test measures allows researchers to control for a range of threats to internal and external validity such as unrelated organisational financial gains or mental health issues (Grant, 2009; Grant, 2012).

A recent inclusion on the list of measurable coaching outcomes is Return On Investment (ROI), which involves the coachee examining organisational results and making a judgement call as to how much of the improvement can be attributed to the coaching (McGovern et al., 2001). Many researchers have argued against the use of ROI, given the highly subjective figure being thrust into a formula that ignores many other variables (see De Meuse et al., 2009; Grant, 2012). While it is understandable that those involved in the change sequence should be the closest and most knowledgeable about where their bottom line is affected and how, this form of measurement lacks objectivity. However from the client's point of view, suggesting a potential ROI is often the difference between gaining approval for the coaching project or not. While ROI may not be the most scientific approach to examining effectiveness, practitioner's find it a very useful tool for "proving" their worth to organisations (Parker-Wilkins, 2006). However, ROI is an exceptionally subjective outcome that carries with it a wide range of reliability and validity issues (De Meuse et al., 2009; Grant, 2012). There has been some conjecture amongst researchers around the appropriateness of linking company profits to coaching outcomes (Grant, 2012). The risk of doing so is that coaches will make promises to clients about the outcomes of coaching based on inaccurate data, resulting in damage to its credibility in the market. Therefore objective key performance indicators (KPIs) specific to the individual are recommended. Typically, when individuals rate their performance they tend to overestimate their value in comparison to observer's ratings and concrete data (Crowne & Marlowe, 1960), therefore when measuring performance outcomes objective measures should be taken. This also highlights the need for more performance rating in the coaching literature, as the majority of tools used are related to well-being and self-regulation related topics such as depression, anxiety, and coping. Suggested measures that are relevant to the coach-coachee relationship as well as the organisation funding the program include for example KPIs, staff engagement, and employee turnover.

To further prove the true value of coaching, there is opportunity for researchers to follow the lead of Finn (2007) and Green (2004) by conducting longitudinal explorations. Although sparse, results so far have shown coaching has an incredibly positive, lasting impact on an individual. It should be no surprise that the benefits of coaching remain present with the coachee for months following the final session (Green et al., 2006; Green et al., 2007; Finn, 2007) as improved self-regulation is a key outcome of most coaching programs. Unfortunately, longitudinal studies to date have only measured experimental groups for the longitudinal component. By taking measures from the experimental and control groups over multiple time points, future researchers could extend these data to find some of the lasting impacts that coaching can have. For example, measuring engagement across teams and the wider organisation could provide insight into the top-down flow on effects of coaching and changes in group attitudes.

Conclusion

Overall, coaching resulted in a moderate positive effect for the outcomes included in this study. However the current meta-analysis revealed more about what was missing than what was present. There is a strong empirical claim that coaching improves the well-being of coachees, however the measures employed for performance are not currently meeting a satisfactory empirical standard. The practice of coaching stands to lose its considerable credibility in the commercial world if coaches cannot accurately demonstrate ROI that involves definitive improvements in coachee performance. Theeboom et al (2014) asked, 'how does it work?'–clearly it can be argued that coaching improves the well-being of the coachee, which in turn is seen as favourable by the organisation. However the question to be asked now is, 'what is the value?' more specifically, 'how do we measure the value?' Current research around return on investment from a commercial point of view is limited. Either psychologists need to become more commercial or coaches in general need to take on more statistical rigour to prove their worth to organisations.

References

Bax, L., Yu, L. M., Ikeda, N., & Moons, K. G. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology*, *7*(1), 40.

Berger, L. A., & Berger, D. R. (2004). The talent management handbook. *New York*.

Bolch, M. (2001). Proactive coaching despite costs of thousands of dollars a day, executive coaching-has emerged as one of the hottest trends in the business world. *Training-New York Then Minneapolis-*, *38*(5), 58-66.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.

*Bozer, G., & Sarros, J. (2012). Examining the effectiveness of executive coaching on coachees' performance in the Israeli context. *International Journal of Evidence Based Coaching and Mentoring [E]*, *10*(1), 14-32.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of research.* Boston: Houghton Mifflin Company.

Cochran, W.G. (1954). The combination of estimates from different experiments. *Biometrics, 10:*101–129.

Cohen, S. S. (1988). *Practical statistics*. E. Arnold.

Coutu, D., & Kauffman, C. (2009). What can coaches do for you. *Harvard Business Review*, *87*(1), 91-97.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(3), 245-265.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

De Meuse, K., Dai, G., & Lee, R. (2009). Does executive coaching work: A meta analysis study. *Coaching: An International Journal of Theory, Practice and Research*, *2*(2), 117-134.

Deviney, D. E. (1994). The effects of coaching using multiple rater feedback to change supervisor behavior. *Dissertation Abstracts International Section A, 55,* 114.

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed-and random-effects methods. *Psychological methods*, *6*(2), 161.

*Finn, F. A. (2007). *Leadership development through executive coaching: The effects on leaders' psychological states and transformational leadership behaviour.* Unpublished doctoral thesis, Queensland University of Technology.

*Grant, A. M. (2001). *Towards a psychology of coaching: The impact of coaching on metacognition, mental health and goal attainment.* Unpublished doctoral thesis, Macquarie University.

Grant, A. M. (2003). The impact of life coaching on goal-attainment, metacognition and mental health. *Social Behavior and Personality, 31*, 253-264.

Grant, A. M. (2008). Personal life coaching for coaches-in-training enhances goal attainment, insight and learning. *Coaching: An International Journal of Theory, Research and Practice, 1*(1), 54-70.

Grant, A. M. (2009). Workplace, executive and life coaching: An annotated bibliography from the behavioural science and business literature. *Coaching Psychology Unit, University of Sydney, Australia*.

Grant, A. M. (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.

Grant, A. M., & Cavanagh, M. J. (2004). Toward a profession of coaching: Sixty-five years of progress and challenges for the future. *International Journal of Evidence Based Coaching and Mentoring*, *2*(1), 1-16.

*Grant, A.M., Curtayne, L., & Burton, G. (2009). Executive Coaching enhances goal attainment, resilience and workplace well-being: A randomized controlled study. *The Journal of Positive Psychology, 4*, 396-407.

*Grant, A.M., Green, L.S., & Rynsaardt, J. (2010). Developmental Coaching for high school teachers: Executive coaching goes to school. *Consulting Psychology Journal: Practice and Research. 3*, 151-168.

Grant, A. M., Passmore, J., Cavanagh, M. J., & Parker, H. M. (2010). The State of play in coaching today: A comprehensive review of the field. *International Review of Industrial and Organizational Psychology*, *25*(1), 125-167.

*Green, S. (2004). *The efficacy of group-based life coaching: A controlled trial.* Unpublished doctoral thesis, University of Wollongong.

*Green, L. S., Grant, A. M., & Rynsaardt, J. (2007). Evidence based life coaching for senior high school students: Building hardiness and hope. *International Coaching Psychology Review, 2*, 24 – 32.

*Green, L. S., Oades, L. G., & Grant, A. M. (2006). Cognitive behavioural, solution-focused life coaching: Enhancing goal striving, well-being and hope. *Journal of Positive Psychology, 1*, 142 – 149.

Greif, S. (2007). Advances in research on coaching outcomes. *International Coaching Psychology Review*, *2*(3), 222-249.

Hay Group. (2002). *The Future of Executive Coaching,* London: Hay Group.

Hedges, L. V. (1994). Fixed effects models. *The handbook of research synthesis*, 285-299.

Hedges, L. V. (2008). What are effect sizes and why do we need them? *Child Development Perspectives, 2*(3), 167-171.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558.

International Coach Federation. (2012). Retrieved May, 16, 2012 from, http://www.coachfederation.org/about-icf/overview/

Kluger, A., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback theory. *Psychological Bulletin* 199.2: 254–284.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*.

Leadership Management Australia. (2006). *The L.E.A.D. Survey 2005/6*. Melbourne: Leadership Management Australia.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335-343.

McGovern, J., Lindemann, M., Vergara, M., Murphy, S., Barker, L., & Warrenfeltz, R. (2001). Maximizing the impact of executive coaching. *Manchester Review*, *6*(1), 2001.

Moen, F., & Federici, R. A. (2012). The effect of external executive coaching and coaching-based leadership on need satisfaction. *Organization Development Journal*, *30*(3).

*O'Connor, S., & Cavanagh, M. (2012). *The coaching ripple effect: The individual and systemic level influence of leadership development*. Unpublished doctoral thesis, Coaching Psychology Unit, University of Sydney, Sydney, Australia.

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

Poepsel, M.A. (2011). *The Impact of an Online Evidence-Based Coaching Program on Goal Striving, Subjective Well-Being, and Level of Hope*. Unpublished doctoral thesis, Capella University.

*Richardson, T. M. (2010). *Solution-Focused Brief Coaching as an Executive Coaching Intervention: A Quasi-experimental Study*. Unpublished doctoral thesis, University of Phoenix.

Schmidt, F., & Hunter, J. E. (1995). The impact of data-analysis methods on cumulative research knowledge statistical significance testing, confidence intervals, and meta-analysis. *Evaluation & the Health Professions*, *18*(4), 408-427.

Sherman, S., & Freas, A. (2004). The wild west of executive coaching. *Harvard Business Review*, *82*(11), 82-93.

Smither, J. W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personnel Psychology*, *56*(1), 23-44.

*Spence, G. B. (2006). *New directions in the psychology of coaching: the integration of mindfulness training into evidence-based coaching practice.* Unpublished doctoral thesis, University of Sydney.

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

Spence, G.B., Cavanagh, M.J., & Grant, A.M. (2008). The integration of mindfulness training and health coaching: An exploratory study. *Coaching: An International Journal of Theory, Research and Practice, 1,* 145-163.

Spence, G. B., & Grant, A. M. (2007). Professional and peer life coaching and the enhancement of goal striving and well-being: An exploratory study. *Journal of Positive Psychology, 2*, 185 – 194.

Spielberger, C. D. (1980). *Test anxiety inventory*. John Wiley & Sons

Sterne, J., Egger, M., & Moher, D. (2008). Addressing reporting biases. In J. P. Higgins (Ed.), *Cochrane handbook for systematic reviews of interventions* (pp. 297-334). Chichester: Wiley-Blackwell.

Stober, D. R., & Grant, A. M. (Eds.). (2006). *Evidence based coaching handbook: Putting best practices to work for your clients*. John Wiley & Sons.

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.

Tobias, L. L. (1996). Coaching executives. *Consulting Psychology Journal: Practice and Research, 48*(2), 87.

Wasylyshyn, K.M. (2003). *Executive Coaching – An Outcome Study.* Published online. Retrieved www.karolwasylyshyn.com/pdf/executive_coaching.pdf

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

*Indicates studies included in the present meta-analysis

Daniel Burt holds a Master of Applied Psychology, Organisational Psychology from Murdoch University and works in the Organisation and People Development Department of the Office of the Director of Public Prosecutions for Western Australia.

Zenobia Talati holds a Doctor of Philosophy in Social Psychology and a Master of Industrial and Organisational Psychology from the University of Western Australia. She is currently a Postdoctoral Research Associate in the School of Psychology and Speech Pathology at Curtin University.