Paper

Who sees change after leadership coaching? An analysis of impact by rater level and self-other alignment on multi-source feedback

Doug MacKie

Objectives: *The objective of this research was to investigate the use of multi-source feedback in assessing the effectiveness of a strength-based coaching methodology in enhancing elements of the full range leadership model. It also investigated the effects of self-other rater alignment on leadership outcomes after coaching.*

Design: *A between-subject non-equivalent control group design was used to explore the impact of strength-based coaching on transformational leadership behaviours measured in a 360-degree feedback process. Thirty-one executives and senior managers from a large not-for-profit organisation were non-randomly assigned to either a coaching or waitlist cohort.*

Methods: *The coaching cohort received six sessions of leadership coaching involving feedback on leadership and strengths, goal setting and strengths development. After six sessions of coaching over three months, cohorts then switched roles.*

Results: *The results showed that participants experienced statistically significant increases in their transformational leadership behaviour after coaching and this difference was perceived differentially at all levels within the organisation but not by the participants themselves. Raters at higher levels in the organisation were the most sensitive to change. The results also showed that self-other rater alignment was a significant factor in self-ratings of change over time with those participants who initially over-rated themselves, reducing their ratings over time as a consequence.*

Conclusion: *The results suggest that changes in coachee transformational leadership behaviour after leadership coaching are perceived differentially by rater level within an organisation and that self-other rater alignment is an important moderator of self-ratings over time.*

Keywords: *leadership coaching; strength-based; rater level; multi-source feedback; self-other alignment.*

EVIDENCE IS GROWING for the effectiveness of executive coaching in organisations and yet much of the research shows an over-reliance of self-report measures rather than investigating the impact more broadly on managers, peers and direct reports within the organisation, (MacKie, 2014; Page & de Haan, 2014). Multi-source feedback (MSF) or 360-degree feedback provides the opportunity to examine the impact of executive coaching more broadly in the organisation and to extend the analysis of impact beyond the level of self-report. Multi-source feedback or 360 methodologies are near ubiquitous in leadership development programmemes including coaching approaches and yet they are primarily used for assessment and awareness raising purposes rather than as formative outcome criteria to assess change after leadership development interventions (Kochanowski, Seifert & Yukl, 2010; Nowack & Mashihi, 2012).

There are several reasons why MSF within a leadership framework are desirable dependent variables in leadership coaching outcome studies. Self-reports are an unreliable indicator of change, especially when the

rater is least skilled in that area, (Kruger & Dunning, 1999) and are prone to leniency bias (Fleenor, Smither, Atwater et al., 2010). Secondly leadership coaching by definition, requires an impact beyond the self-report of the coachee as the purpose of modifying their behaviour is to have a more transformative impact on those around them in an organisation (Kochanowski, Seifert & Yukl, 2010). Indeed there is significant evidence linking perceived changes in leader behaviour with enhanced engagement and discretionary effort in followers, resulting in enhanced business-unit outcomes, (Avolio, 2011; Harter, Schmidt & Hayes, 2002). Prior research has also found that significant changes in transformational leadership after coaching have been perceived by other raters with the same organisation but these studies have not reported changes by rater level, (Cerni, Curtis & Colmar, 2010; O'Connor & Cavanagh, 2013). Transformational leadership is one of the most researched leadership theories over the last 30 years and has established significant correlations between increases in transformational leadership and objective performance outcomes including financial performance, job satisfaction, follower satisfaction, and organisational commitment (Avolio, 2011).

Finally, multi-source leadership ratings can provide a reliable and valid outcome measure that can be compared across studies, a key criterion if different coaching methodologies are to be compared and contrasted, (MacKie, 2014). Evidence for the validity of other-ratings has been demonstrated in prior research. Atkins and Wood (2002) used assessment centre ratings as the objective and independent performance criteria and found that other-raters derived from MSF significantly predicted performance on the assessment centre, as did line manager ratings alone. However, it is important to remember that in leadership assessment, as in so many job performance criteria, there is no 'objective' measure of performance primarily due to criterion deficiency, that is aspects of job performance like team work are not easily amenable to objective measurement, (Murphy, 2008).

It is important to recognise that even without subsequent leadership coaching, MSF is an intervention in itself, especially if formally debriefed with the participant (Neiminen et al., 2013). In a meta-analysis of 24 longitudinal studies, small but significant effect sizes in performance were found after MSF as observed by supervisors and the participant's direct reports (d=0.15), peers (d=0.05) but not self-ratings (d=–0.04), (Smither, London & Reilly, 2005). There is also some evidence to suggest that raters at different levels focus on different aspects of the leader with supervisor ratings being more closely correlated with external performance criteria (Atkins & Wood, 2002) whilst direct reports (those reporting directly into the participant), focus on more interpersonal and relational criteria (Nowack, 2009).

Another critical issue in the application of MSF to the evaluation of leadership coaching interventions is the issue of self-other agreement (SOA) (Fleenor et al., 2010). Given the challenges of self-report data, it is not surprising to find that there are typically modest correlations between self and other ratings in the existing literature, (Fleenor et al., 2010, Nowack & Mashihi, 2012). However, within these discrepancies, there are a variety of interesting sub-groups. Individuals who overrate themselves compared to other raters are seen to be potentially at risk of derailing (ie. failing to reach their career apogee) and may be less receptive to feedback, (Woo et al., 2008) but the gap between self and others can also act as a motivator to change, (Atwater & Brett, 2005). Individuals who under-rate themselves when compared to others are potentially self-critical and perfectionistic and can be demotivated by the realisation that others perceive them to be more effective than they see themselves, (Nowack & Mashihi, 2012). These somewhat contradictory findings in the SOA literature suggest the impact of misalignment of self-other ratings on coach-

ing outcomes would be a useful hypothesis to explore further.

Finally there is a developing body evidence of that has specifically examined perceived changes in MSF after coaching interventions. Luthans and Petersen (2003) investigated the effects of feedback plus coaching on a group of 27 managers and 67 employees. The study specifically aimed to reduce the self-other discrepancies in MSF. At Time 1 there were significant differences in the self-other ratings (with self being higher) on behavioural and interpersonal competency and personal responsibility. However, these discrepancies had all disappeared at Time 2 after the coaching process. Interestingly the self-ratings had not diminished over time but the other ratings had caught up at Time 2. Unfortunately all other raters were combined into one category so it was not possible to ascertain the differences by organisational level. The results also suggested an improvement in manager and employee satisfaction and commitment at Time 2.

Smither et al. (2003) examined whether coaching could improve the effect of 360-degree feedback in enhancing performance. Of 1202 senior managers who received 360-degree feedback, 404 were selected for subsequent coaching. Those who participated in coaching were reported to set more specific goals, solicit more ideas and improve more in terms of others' ratings. However despite some positive changes in goal setting and performance (Effect size, d=0.17) as measured by supervisor and direct report ratings (self reports were not analysed and peer ratings were not significant) in a repeat 360 feedback process, only 30 per cent of the original participants participated in the coaching and the selection criteria for their inclusion was mixed (some were required to participate by their managers) making the results prone to selection effects and difficult to generalise.

Kochanowski, Seifert and Yukl (2010) investigated the effective of a feedback workshop plus coaching on the influencing skills of managers. Thirty managers were randomly assigned to either workshop alone or workshop plus telephone coaching. Feedback was based on the manager's boss and at least three direct reports. The results showed that the coaching group demonstrated a significant increase on the control group in only one of the four influencing tactics assessed (collaboration). However only the subordinate data was used in the comparative analysis so the impact on different raters levels (e.g. boss, peer or self) remained unknown. In a study of 469 middle managers, Neiman et al. (2013) used a quasi-experimental design to compare the impact of MSF alone and MSF plus five sessions of executive coaching. The results suggested that while both groups improved equally when rated by direct reports, peers and supervisors, only managers who had received the coaching improved on self-ratings of leadership behaviour and effectiveness.

In conclusion, whilst there is growing evidence for the effectiveness of executive coaching in organisations, the level of that impact remains unclear. MSF provides an opportunity to extend the evaluation of executive coaching beyond the reliance on self-report and assess its broader impact in the organisation. In addition the issue of self-other alignment has potentially significant implications for coaching evaluation given that positive discrepancies may suppress the coaching effect as the coachees adjust their ratings to align more with other raters. Multi-rater assessment based on reliable and valid leadership constructs provides the opportunity to track the impact of executive coaching as its effects cascade through the organisation.

Rationale and aims

The limited number of studies that have examined the impact of leadership coaching interventions by level and considered the impact of SOA have demonstrated inconsistent results and drawn different inferences making conclusions about effectiveness difficult to generalise. This study aims to address

these issues in coaching research by specifically investigating the impact of perceived coaching outcomes by the level of raters within an organisation and investigating the impact on outcomes of self-other alignment (SOA).

The dependent variable used to assess outcomes was the full range leadership model (FRLM) that includes transformational, transactional and laissez-faire elements of leadership. This leadership outcome provides 360-degree feedback on changes in leadership behaviour throughout the organisation and moves the assessment of coaching outcomes beyond the reliance on self-report measures. By focusing on a specific strength-based coaching methodology, using a reliable and valid measure of transactional and transformational leadership as the dependent variable and assessing outcomes by way of a 360-degree feedback methodology, this study aims investigate the level at which change in leadership behaviours after coaching is perceived within the organisation and the impact of SOA on leadership outcome ratings after coaching.

Hypotheses

The following specific research questions will be addressed in an attempt to address the aims of the study.

1. Changes in transformational leadership behaviour observed after participation in a coaching process will be perceived differentially throughout the organisation. The perception of change in leadership behaviours will vary by the level of the rater with line managers and direct reports reporting most change followed by peers and self-reports.
2. Self-other agreement at Time 1 will impact subsequent self-ratings of transformational leadership. Participants who over-estimate their MSF ratings as compared to all others will show a tendency to reduce their ratings over time. Participants who underestimate their ratings compared to all others will increase their ratings over time.

Method

Participants

A total of 31 senior managers (14 male, 17 female) were recruited from the same organisation in the not-for-profit (NFP) sector. They were all senior managers and leaders in the Australian arm of a multi-national not-for-profit (NFP) organisation. The average age was 45 years (range 31 to 62 years). This represented all available senior mangers from the top two levels in the organisation and included the executive director and the leadership team. A total of 37 individuals were invited to participate but six were unable to participate due to overseas postings and maternity leave. Having managerial responsibility for a number of direct reports was a prerequisite of participating in the study. The participants were then divided into two groups – the coaching first group (Cohort 1), and the waitlist first group (Cohort 2). The process of group allocation was not random as it depended on the availability of the participants and the preferences of the organisation. All participants gave their written informed consent to participate in the study.

Raters

Raters were all drawn from the same organisation as the participants. They were a mixture of line managers, peers, direct reports and others in the organisation who did not fit into the first three categories. Each participant had an average of 9.8, 9.7 and 9.6 raters at the three time points respectively. Rater consistency across time was high with 92.5 per cent and 88.8 per cent of the original raters also responding at Time 2 and Time 3. There was no significant difference between the ratings of original and original plus new raters at Time 2 and 3 on transformational leadership, hence the full compliment of raters was used in the analysis. (See MacKie, 2014, for a full description of the rater consistency data.)

Research design

The study utilised a non-equivalent control group design with two cohorts; a Coaching

first group (Cohort 1) and a Waitlist first group (Cohort 2). While cohort 1 was engaged in the coaching, Cohort 2 acted as the control group. Cohorts then switch roles at the mid-point (Time 2). However, because Cohort 1 had had the coaching intervention at this stage, it was not able to act as an independent control group for Cohort 2. Each participant received six sessions (nine hours) of strength-based leadership coaching. The dependent variable was the multi-source feedback on participant's transformational leadership behaviour provided by the MLQ 360. Each participant had an average of 9.8 raters from above, peer, direct report and other levels within the organisation.

Coaches

A total of 11 coaches provided their services pro-bono for the research. They were highly experienced practitioners who were mainly recruited from the local executive education department of a prestigious business school and had been preselected for both psychological mindedness and business acumen. All coaches were self-employed practitioners who earned a significant part of their income from providing executive coaching services to corporate entities. On average they had 12 years of experience providing executive coaching in organisations and had been working in organisations for an average of 28 years. The majority (70 per cent) were qualified at Masters level or above and were registered practising psychologists. Each coach was trained in the author's strength-based methodology by way of a half-day training programmeme. This process described the underlying rationale for strength-based approaches to leadership and provided a structured strength-based coaching manual for the coach to follow. The induction particularly focused on the identification of strengths through interview data, MLQ 360-degree feedback and the Realise 2 inventory (Linely & Stoker, 2012). The Realise 2 four quadrant model was also used to give the coaches a format for setting strength-based development goals. The induction also provided the format for strengths development through the sessional rating of strengths awareness, alignment, pairing and utilisation. Each coach provided leadership coaching to between one to two participants per cohort.

Procedure

Strength-based protocol. Each coachee received six 90-minute coaching sessions that followed a format articulated in their coaching manual. Initially coaches began with a strength-based interview followed by feedback for the coachee on their MLQ 360 report (Bass & Avolio, 1997) and Realise 2 Inventory (Linley, Willars & Biswas-Diener, 2010). The strength-based interview focused on their peak experiences and what energised them about their work. The Realise 2 questionnaire provided feedback on what energised them, where they felt competent and where they had the opportunity to apply their strengths. This led to structured feedback on their realised strengths (those that were known and utilised), unrealised strengths (those that were known but underutilised), learned behaviours (those that were competent but not energising) and weaknesses (where both competence and energy were low). The MLQ 360 provided qualitative and quantitative multi-rater feedback on their scores on the full range leadership model (FRLM) that included transformational, transactional and laissez-faire leadership styles. Coachees were then required to select three goals they would like to focus on during the coaching; a realised strength, an unrealised strength and a learned behaviour or weakness. These goals were focused on issues that the coachee was motivated in addressing and also that had relevance for the business. Coachees then tracked their progress on these goals for the remaining five sessions and committed to actions designed to help their goal attainment. Coachees also tracked their progress on a sessional basis by reflecting on and rating their strength awareness, alignment, pairing and utilisation in their coaching

manual. This study also utilised a measure of protocol and manual adherence to investigate the link between adherence to the strength-based approach and subsequent changes in transformational leadership behaviour. (See MacKie, 2014 for further elaboration.)

Measures

Each participant received the *Multi-Factor Leadership Questionnaire*. The MLQ 360 (Bass & Avolio, 1997) is a 49-item questionnaire that measures nine elements of the full range leadership model (FRLM) namely idealised influence attributes (e.g. Display a sense of power and confidence), idealised influence behaviour (e.g. Talk about my most important values and beliefs), inspirational motivation (e.g. Articulate a compelling vision of the future), intellectual stimulation (e.g. Seek different perspectives when solving problems), individualised consideration (e.g. Help others to develop their strengths), contingent reward (e.g. Provide others with assistance in exchange for their efforts), management by exception active (e.g. Keep track of all mistakes), management by exception passive (e.g. Fail to interfere until things become serious) and laissez-faire (e.g. Avoid making decisions). The inventory also has three measures of leadership outcomes; extra effort (e.g. Heighten others' desire to succeed), effectiveness (e.g. Lead a group that is effective) and satisfaction (e.g. Work with others in a satisfactory way) (Bass & Avolio, 1997). The MLQ360 measures all items on a five-point Likert scale from 'not at all' to 'frequently if not always'. Cronbach's alpha for the main transformational leadership factor has been reported as 0.85 (Antonakis et al., 2003) and criterion validities vary for satisfaction (0.71), effectiveness (0.64) and performance (0.27) (Judge & Piccolo, 2004).

Data analysis

Repeated measures ANOVAs (Tabachnick & Fidell, 2007) were calculated using SPSS that allowed analysis of both within-subject changes in the dependent variable (DV) over time and between group differences in terms of rate of change on the DV. The five transformational leadership sub-scales as measured by all raters, were aggregated into one combined transformational leadership score, the MLQ 5I (Bass & Avolio, 1997) to provide an overall index of change. These ratings were then divided into mean self-ratings for average transformational leadership and mean other-ratings for composite transformational leadership scores by time to examine the impact of SOA on outcomes.

Results

Hypothesis 1: Changes in transformational leadership behaviour observed after participation in a coaching process will be perceived differentially throughout the organisation. The perception of change in leadership behaviours will vary by the level of the rater with line managers and direct reports reporting most change followed by peers and self-reports.

There were no significant differences between the two groups at Time 1 on the MLQ 5I composite score at any rater level. The key component of this hypothesis was that changes in transformational leadership would be observed differentially beyond the level of self-report. Consequently an analysis by level of rater was conducted to see who observes the changes in participant leadership behaviour and whether the organisational level of the observing rater is a significant factor in observing changes in leadership behaviour.

A repeated measures ANOVA was performed to examine the impact on coaching on mean transformational leadership scores over time by rater level. Table 1 clearly illustrates that in both cohorts the greatest significant change and effect size was achieved by the higher-level raters. This suggests that those raters working above the participant in the organisation, for example, their line manager, were seeing the greatest change in the participants in terms of transformational leadership behaviour after their leadership

coaching. For both cohorts, the peer and direct report level perceived significant change in participant's transformational leadership behaviour at the $p<0.05$ level. For the lower level, both groups reported significantly higher ratings on transformational leadership after participant coaching (Table 1).

It is important to note that there were no significant positive changes over time in the participants own perceptions of their transformational leadership behaviour although there was a positive but non-significant trend in Cohort 2. In Cohort 1 the self-ratings actually declined over time despite all other-raters reporting a positive increase in transformational leadership behaviour. This may be due to the lack of SOA at T1 for Cohort 1, a hypothesis more fully explored in the next section. An analysis of the five individual components of transformational leadership revealed that the only significant decline in self-ratings was in inspirational motivation (IM) in Cohort 1 ($F(1,29)=4.781$, $p=0.040$, partial $\eta^2=0.179$).

Hypothesis 2: Self-other agreement at Time 1 will impact subsequent self-ratings of transformational leadership. Participants who over-estimate their MSF ratings as compared to all others will show a tendency to reduce their ratings over time. Participants who underestimate their ratings compared to all others will increase their rating over time.

In order to examine the impact of self-other alignment on outcomes, a mean other rater composite score was calculated. The results in Table 2 demonstrate that for C1, there was a significant difference in self-other ratings at Time 1 with participants significantly over-rating themselves compared to all other raters. This discrepancy then disappeared at T2 when they appeared much more aligned and then reappeared in the other direction at Time 3 with participants continuing to decrease their self-ratings even after the coaching intervention. For Cohort 2, there was no initial misalignment between self-other ratings and both ratings showed good alignment over the three time points.

Figure 1 shows that C1 participants began with a significantly higher rating than the combined all other rater group. This discrepancy then reduces at Time 2 possibly as a function of increased awareness of the discrepancy during the feedback process. This may explain why the participant scores have decreased after coaching while all other raters level scores have increased. This elevated self-rating was only apparent in Cohort 1 and was not apparent in Cohort 2. Receiving the feedback at Time 1 that all other raters view their MLQ scores at a lower level, appears to have driven down the subsequent MLQ self–ratings even during the coaching process where all other rater levels are reporting an increase in transformational leadership behaviour. At Time 3, the Self-other discrepancy for Cohort 1 has gone the other way with participants significantly underestimating their transformational leadership scores when compared to all other raters.

Figure 1 shows that in contrast to Cohort1, Cohort 2 began with participant MLQ ratings much more aligned to all other raters. Their ratings do not significantly change between T1 and T2 as they have not yet had the MLQ feedback that is embedded in the coaching. At Time 3 after the completion of the coaching, their mean MLQ5I transformational leadership scores remain aligned with all other raters. The results from Cohort 2 suggest that beginning the coaching process with a strong SOA is crucial in maintaining this over time. Despite this alignment, however, the self-ratings in Cohort 2 did not show significant change over time unlike the combined all-other transformational leadership ratings.

Discussion

The results demonstrated three significant findings in the utlilisation of MSF to evaluate outcomes in leadership coaching. Firstly there was no significant change in the participants' ratings of transformational leadership after the coaching intervention. Secondly the change in other-ratings of

Table 1: Changes in Mean Transformational MLQ rater scores for Cohort 1 and 2 across Time 1 and Time 3 by rater level.

	Time 1		Time 3					
MLQ rater level	*M*	*SD*	*M*	*SD*	*F*	*df*	*p*	*Partial* η^2
Cohort 1 Coaching First								
Self (*N*=14)	2.92	0.37	2.72	0.41	1.503	1,24	.233	.064
Higher (*N*=16)	2.64	0.72	3.17	0.60	4.694	1,30	.039	.144
Peer (*N*=49)	2.55	0.66	2.92	0.63	5.666	1,76	.020	.071
Lower (*N*=56)	2.61	0.75	2.97	0.48	6.452	1,90	.013	.068
Other (*N*=9)	2.69	0.94	2.80	0.61	0.165	1,17	.795	.005
Cohort 2 Waitlist First								
Self (*N*=17)	2.90	0.46	3.05	0.46	0.761	1,31	.390	.026
Higher (*N*=22)	2.81	0.55	3.35	0.54	8.685	1,37	.006	.199
Peer (*N*=51)	2.70	0.62	3.10	0.47	12.937	1,100	.001	.117
Lower (*N*=67)	2.80	0.78	3.11	0.54	5.691	1,117	.019	.047
Other (*N*=12)	2.57	0.93	2.87		0.094	1,13	.765	.008

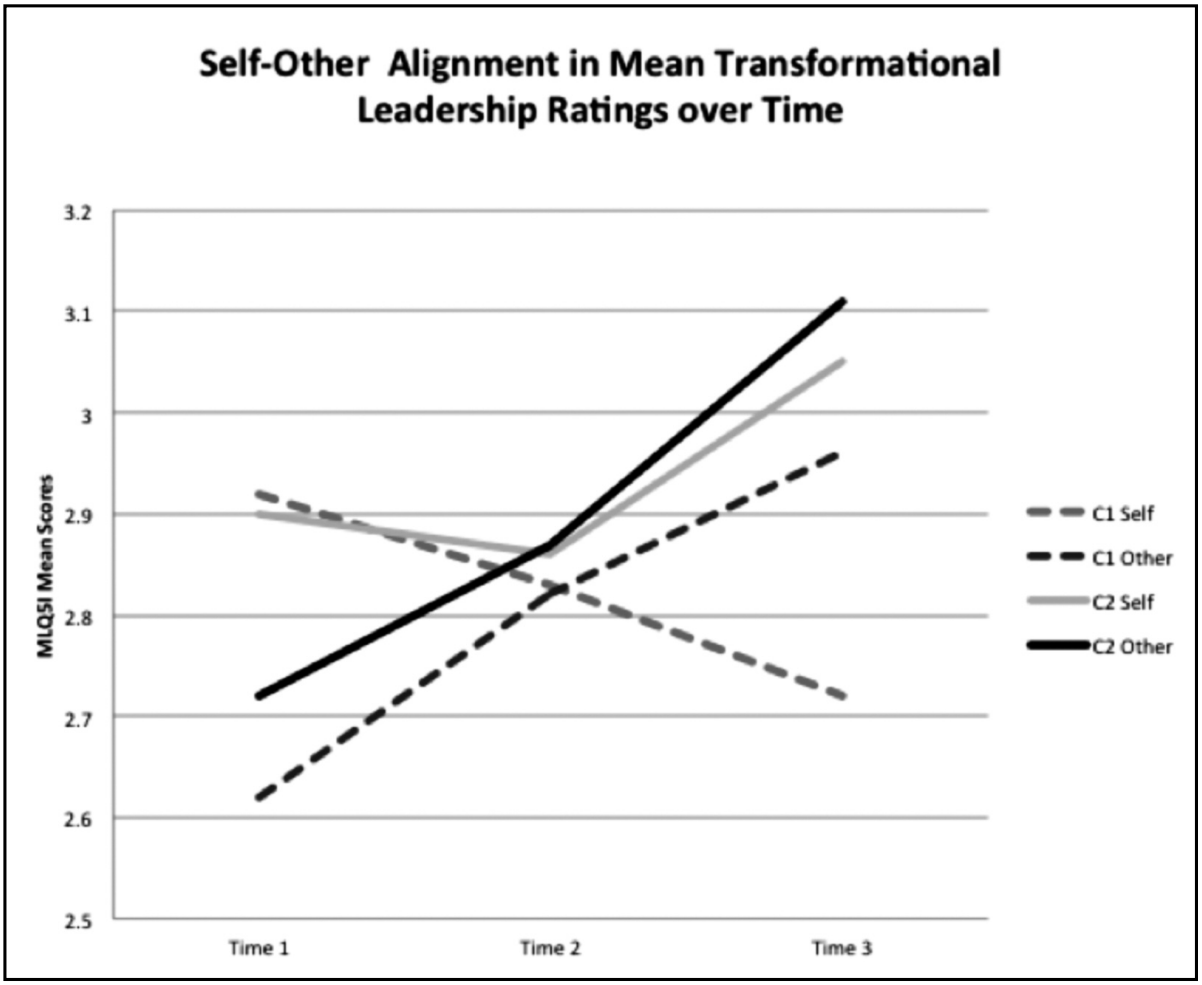
Note: Repeated Measures ANOVA Within Group Comparison of MLQ5I means between Waitlist first group and Coaching first group at Time 1 and Time 3 by rater level. Higher=line manager and lower=direct reports.

Table 2: Self-Other Alignment in Mean Transformational MLQ rater scores for Cohort 1 and 2 across three time points.

	Self		Other					
MLQ rater level	*M*	*SD*	*M*	*SD*	*F*	*df*	*p*	*Partial* η^2
Cohort 1 Coaching First								
Time 1	2.93	0.38	2.60	0.30	15.30	1,14	.002	.541
Time 2	2.83	0.43	2.82	0.32	0.02	1,13	.883	.002
Time 3	2.72	0.41	2.98	0.25	10.66	1,10	.010	.542
Cohort 2 Waitlist First								
Time 1	2.90	0.47	2.75	0.37	1.17	1,17	.294	.069
Time 2	2.86	0.51	2.87	0.43	1.283	1,16	.941	.000
Time 3	3.05	0.46	3.12	0.26	0.33	1,13	.135	.025

Note: Repeated Measures ANOVA Within Group Comparison of MLQ5I means between Waitlist first group and Coaching first group at Time 1, Time 2 and Time 3 by rater level. Self=participant, other=all other raters combined.

Figure 1: Self-Other Alignment in Mean MLQ5I Transformational Leadership scores over time for Cohorts 1 and 2.



transformational leadership scores after leadership are differentially perceived with those higher in the organisation being most sensitive to change. Finally participants who initially overestimate their leadership scores when compared to all other raters, appear to subsequently reduce their scores in an attempt to realign with other raters. This effect may be partially responsible for the lack of significant difference in the self-ratings of participants over time.

The lack of changes in self-ratings of leadership behaviour after leadership coaching is unusual and at odds with many of the previous findings in this area (Grant et al., 2010; Theeboom et al., 2014). There was a non-significant decline in scores over time for Cohort 1 and a small non-significant increase in time for Cohort 2. Their self-ratings on transformational leadership remained very consistent between time one and two while they were acting as the control group for Cohort 1. After the coaching had been received at Time 3, their self-ratings did increase in line with all other levels rating their behaviour. Given that the majority of outcome studies employ only self-report measures, it is interesting to speculate how many studies may have shown a significant other-rating change had that data been available. This finding also confirms that self-ratings alone may be an unreliable indicator of change (Kruger & Dunning, 1993) and prone to under-estimation (Fleenor et al., 2010). There is related evidence to suggest that 360 feedback can have the effect of

lowering subsequent levels of self-rating as the participant's awareness is raised about how others view them but only if participants initially overrate their leadership abilities (Atwater et al., 2000). Other researchers have confirmed that when individuals overrate their leadership behaviour, subsequent ratings can decrease as a function of greater insight and feedback (Luthans & Petersen, 2003). This suggests that the other-rater alignment effect is more powerful that the self-perceived changes after coaching.

However, there were significant differences in how raters at different levels viewed the changes in coachee leadership behaviour over time. There were also some between cohort differences of note that are worth exploring. Cohort 1 (Coaching first) began with a higher self-rating compared with their manager, peers, direct reports and other raters. As outlined above, the self-ratings of transformational leadership behaviour came down after the coaching whilst all other levels of raters increased their ratings. The other finding of note in Cohort 1 was that the change in rater's responses over time was differentially perceived with change most apparent within raters at the higher organisational level. Both peers and direct reports saw significant positive changes over time in the levels of participant transformational leadership but the effect size was lower than in the higher level. This is an unusual finding as previous research has suggested that direct reports are the most sensitive to change both for their proximity to the participant and because their data is based on multiple rather than single observations (Atkins & Wood, 2002). An alternative explanation is that different rater levels are rating different qualities in the participant with higher raters rating performance criteria whilst direct reports are rating relational factors (Nowack, 2009).

The second cohort (waitlist first) did not have such an obvious discrepancy between self and other raters on the MLQ at Time 1. Their self-ratings on transformational leadership remained very consistent between time one and two while they were acting as the control group for Cohort 1. Again in Cohort 2, the line manager raters showed the greatest effect size in their ratings of changes in transformational leadership in the participants over time. Given that both Cohorts demonstrated that the line manger raters saw the greatest amount of change over time, this seems to be a reliable finding. As almost all the managers were also participants in the coaching research, they could be especially attuned to the type of changes in transformational leadership behaviour that the participants were being rated on. Peers and direct reports also reported significant levels of change in transformational leadership but again the effect size was not as great as that of the higher group. Given that self-ratings are prone to a variety of self-serving biases that can both promote an inflated sense of self-performance and restrict access to corrective feedback (Dunning et al., 2003), this further emphasises the importance of the trends in the other rater data.

The third critical finding is that self-other misalignment in ratings at Time 1 (in this case overestimation in Cohort 1) appears to trigger an attempt by the participant to realign their scores with all others. This effect seems more powerful than the coaching effect and has the impact of driving down self-ratings over time. This is consistent with a general trend in multi-source feedback that participant self-ratings become more accurate (that is more aligned with the ratings of others) over time as their awareness of the ratings of others increases, (Atwater et al., 2000; Atwater, Brett & Charles, 2007). However, in this case for Cohort 1 they are most aligned with other raters at Time 2 and then diverge in the opposite direction at Time 3 as they continue to underestimate their scores when compared to all others, possibly due to overcompensating for their initial over-estimation at Time 1.

Practical implications

There are a number of practical implications that can be derived from this study. Firstly it

again questions the veracity of self-report and strongly suggests future coaching outcome studies should employ a multi-rater format to assess change. Secondly it suggests that change is perceived differentially by level within the organisation with those higher in the organisation being especially sensitive to change, particularly if those raters are also participants in the coaching programme This again confirms the importance of line manager support and awareness of the coaching process and the goals therein. Finally this study suggests that the impact of overestimating self-scores when compared to all others may have the consequence of driving down scores over time and this effect may mask the impact of the coaching process that is perceived by other raters.

Limitation of the study

The study employed a non-equivalent between-subjects controlled design that utilised a control group to assess the impact of a leadership coaching intervention on transformational leadership behaviours. It was not possible to randomly assign subjects to each cohort as the availability of participants as the logistical needs of the organisation took precedence. Despite this non-randomisation, however, there was no significant difference between the two cohorts at Time 1 suggesting the allocation of participants did not unduly influence the study. However, the between-subjects design only allowed the first Cohort to be fully controlled as at Time 2 when the two cohorts crossed over, cohort one had already had the intervention and could no longer function as an independent control group.

Secondly, the participant sample size was relatively small and the loss of 15 per cent of the participants who dropped out during the course of the study, could reduce the generalisability of the study. A larger participant sample size would help to address this issue and permit further investigation of the influence of coachee variables in coaching outcomes. However, it is worth noting that the total number of other raters for Cohort 1 and Cohort 2 were 131 and 152 respectively. Thirdly, the absence of a definitive analytical technique in the literature to assess the impact of SOA, (Fleenor et al., 2010) may have limited further analysis of the impact of this discrepancy on Cohort 1. Finally the fact that the coaching was provided pro bono, could have impacted negatively on the commitment of the coaches to the coaching process. These effect sizes may, therefor,e be an underestimate of the potential changes possible. However, this issue may be counteracted by possible positive effects of the coaches' participation in the author's strength-based coaching methodology induction programme. The fact that this study was conducted in a NFP organisation may limit its generalisability but the organisation did have a standard corporate structure and HR processes and despite being an NFP, there was a very strong focus on financial accountability and evaluating return on investment.

Conclusion

MSF is increasingly utilised in the evaluation of leadership and executive coaching. This study confirms its validity as a critical outcome measure by illustrating that self-report may not always be sensitive to change, change is perceived differentially within the host organisation with different stakeholder groups reporting different perceptions of change over time and that misaligned SOA can subsequently inhibit leadership self-ratings especially in the case of those who initially over-estimate their leadership capacity. It confirms the need for coaching outcome research to focus beyond self-report to include the level at which others perceive change (Barling, 2014). Future coaching research needs to routinely incorporate MSF as an outcome criterion and analyse results by level within the organisation to confirm the novel finding that higher raters are more sensitive to change. The issue of SOA could be further explored with the incorporation of a performance criteria independent of the multi-rater data to test the effects of poor SOA on participant performance.

The Author
Dr Doug MacKie
CSA Consulting,
320 Adelaide Street,
Brisbane, Australia.

Correspondence
Dr Doug MacKie
Email Doug@csaconsulting.biz

References

Antonakis, J., Avolio, B.J. & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14*, 261–295.

Atkins, P.W. & Wood, R.E. (2002). Self-versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programmes. *Personnel Psychology, 55*, 871–904.

Atwater, L.E., Brett, J.F. & Charles, A.C. (2007). Multisource feedback: Lessons learned and implications for practice. *Human Resource Management, 46*(2), 285–307.

Atwater, L.E. & Brett, J.F. (2005). Antecedents and consequences of reactions to developmental 360-feedback. *Journal of Vocational Behaviour, 66*(3), 532–548.

Atwater, L.E., Waldman, D.A., Atwater, D. & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology, 53*, 275–297.

Avolio, B.J. (2011) *Full range leadership development* (2nd ed.). London: Sage.

Barling, J. (2014). *The science of leadership: Lessons from research for organisational leaders*. Oxford: Oxford University Press.

Bass, B.M. & Avolio, B.J. (1997). *Full range leadership development: Manual for the Multifactor Leadership Questionnaire*. Palo Alto, CA: Mind Garden Inc.

Dunning, D., Johnson, K., Ehrlinger, J. & Kruger, J. (2003). Why people fail to recognise their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.

Fleenor, J.W., Smither, J.W., Atwater, L.E., Braddy, P.W. & Sturm, R.E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly, 21*(6), 1005–1034.

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 Organisational Psychology, 25*, 125–167.

Harter, J.K., Schmidt, F.L. & Hayes, T.L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*(2), 268.

Judge, T.A. & Piccolo, R.F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology, 89*, 755–767.

Kochanowski, S., Seifert, C.F. & Yukl, G. (2010). Using coaching to enhance the effects of behavioural feedback to managers. *Journal of Leadership & Organisational Studies, 17*(4), 363–369.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognising one's own incompetence lead to inflated self–assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Linley, A., Willars, J. & Biswas-Diener, R. (2010). *The strengths book*. Coventry: CAPP Press.

Linley, A. & Stoker, H. (2012). *Technical manual and statistical properties for Realise2*. Coventry: Centre of Applied Positive Psychology.

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

MacKie, D.J. (2014). The effectiveness of strength-based executive coaching in enhancing full range leadership development. A controlled study. *Consulting Psychology Journal. Practice and Research, 66*, 118–137.

Murphy, K.R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organisational Psychology, 1*(2), 148–160.

Nieminen, L.R., Smerek, R., Kotrba, L. & Denison, D. (2013). What does an executive coaching intervention add beyond facilitated multi-source feedback? Effects on leader self-ratings and perceived effectiveness. *Human Resource Development Quarterly, 24*(2), 145–176.

Nowack, K.M. & Mashihi, S. (2012). Evidence-based answers to 15 questions about leveraging 360-degree feedback. *Consulting Psychology Journal: Practice and Research, 64*(3), 157.

Nowack, K.M. (2009). Leveraging multi-rater feedback to facilitate successful behavioural change. *Consulting Psychology Journal: Practice and Research, 61*(4), 280.

Page, N. & de Haan, E. (2014). Does executive coaching work? *The Psychologist, 27*(8), 582–586.

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

Smither, J.W., London, M. & Reilly, R.R. (2005). Does performance improve following multi-source feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*(1), 33–66.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (6th ed.). London: Pearson.

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 organisational context. *The Journal of Positive Psychology, 9*(1), 1–18.

Woo, S.E., Woo, S.E., Sims, C.S., Rupp, D.E. & Gibbons, A.M. (2008). Development engagement within and following developmental assessment centers: Considering feedback favorability and self-assessor agreement. *Personnel Psychology, 61*(4), 727–759.