Fairy tales, facts and the future

predicting leadership effectiveness

A common phenomenon and problem in leadership practice concerns undue reliance on popular fads without sufficient consideration given to the validity of these ideas.

Stephen Zaccaro & Zachary Horn, Leadership Quarterly

Overview

Talent management as a discipline is based on two fundamental assumptions:

- we can improve the accuracy of our predictions to forecast who will be successful in future
- we can accelerate the processes for development to improve the success rate of the individuals we identify and select

These assumptions underpin the practices organisations deploy in recruitment and selection, performance management and promotion, career development and succession planning. And if these assumptions are flawed, then we can revert to a time when ambitious employees simply took their chances in selection interviews and then relied on their work performance (or relationship with their boss) to be an indicator of their potential to progress, and their own efforts in self managed development to equip them for future roles.

Elsewhere we've explored the issue of accelerated development and the strategies and tactics to improve processes in leadership development (1). This article examines the challenge of prediction, specifically the accuracy with which we can forecast today who will be effective as tomorrow's leaders.

How well are we doing when we place our bets to play the game of leadership prediction?

Although often caught up in technical - and confusing - arguments between the psychometricians and the assessment specialists, prediction is not a trivial issue. Given a recent period of leadership folly and failure and the associated damaging business consequences, how we select and appoint our current leaders, and the mechanisms we use to identify future leadership, will be key to sustained organisational success. As Warren Bennis famously said: "the quality of all our lives is dependent on the quality of our leadership."

This article began with a client who asked us to conduct a desk top review of the world of leadership assessment and prediction. The focus then is on **leadership**. This analysis may or may not apply to other organisational roles.

The aim: to evaluate the latest round of research to locate the facts within the fairy tales.

In an appeal to humility, it's also worth asking: why against the backdrop of a massive increase in assessment activity, haven't we seen a corresponding leadership renaissance, evident in higher levels of employee trust and engagement, and organisational productivity and innovation? (2) If our systems for the identification and selection of leaders are becoming so much more accurate, why haven't we observed a significant gain in the outcomes of leadership capability?

Are we getting prediction wrong? If so, why?

Even worse, is the leadership assessment industry part of the problem?

And what should we do about it?

Predictive fairy tales

When Lynn Tulip proposes “Saville Consulting Wave increases predictive validity to 0.9.” (3)

The Hay Group, co-developers of the ECI, claim “emotional intelligence accounts for more than 85% of star performance in top leaders.” (4)

Or a researcher finds that “Career Path Appreciation, combined with a personality assessment technique, yielded a validity coefficient of .93.” (5)

Something strange is going on and we find ourselves in a fairy tale world where we measure “once upon a time” and know exactly who will “live happily ever after”.

Apart from the absence of any evidence to substantiate these extravagant claims, it seems difficult to know what is in fact being proposed. If it is being suggested that we have the assessment tools to predict extraordinary success among any leadership group for any outcome, over any time-scale, we now live in a fantasy world of fairy tales.

This article summarises the research based on peer-reviewed studies published in mainstream professional journals. In house proprietary analysis is informative and indicative; it can also be the outcome of selective cherry picking of the positive results and the filing away of the troublesome findings (6). Because **big claims require big evidence**, we focus on validation findings that have undergone the critical scrutiny of professionally well informed peers to differentiate the empirical facts from the fairy tales of marketing hype.

Of course the issue isn’t simply which of the predictor options on offer are more or less valid. “Research proves” might establish which assessment methods are worth considering and which should be ignored. “Research proves” doesn’t however address the practical challenges faced by leadership resourcing professionals: which option will add most predictive power in a specific leadership scenario, and how we should interpret candidate data to make decisions that improve the predictive odds?

Throughout this summary we report predictive power as a validity coefficient, an index indicating the strength of the relationship between the predictor and the outcome, on a scale from 0 (no prediction) to 1.00 (perfect prediction).

This validity coefficient can also be expressed as a percentage of explained variance in effectiveness. A validity coefficient of 0.3, for example, amounts to a forecast of 9% of future effectiveness.

The beginnings of the predictive game

Taking a historical perspective, the problem of prediction is a relatively recent one.

When leadership appointments were made on the basis of military might, political allegiance, tribal loyalty or family membership, we didn't have to worry too much about the identification of future leaders. It has only been in the last 100 years or so that organisations began to look at alternative ways to decide "who rules" and rethink the mechanisms we use to identify, equip and select our leaders.

Progressive business leaders have always thought long and hard about the individuals who should progress within their organisations. Andrew Carnegie, the 19th century "steel king", applied his famous "go the extra mile" (7) rule in his review of managerial prospects. Here Carnegie recognised the difference that exceptionally talented individuals could make to the success of his firm and looked for that energetic curiosity he believed indicated something important about future effectiveness.

"People are developed the same way gold is mined. Several tons of dirt must be moved to get an ounce of gold. But you don't go into the mine looking for dirt, you go in looking for gold." The "glitter of the gold" was evident in the extent to which individuals put in effort over and above the call of duty to perform more and better service.

It wasn't a bad talent strategy at the time. Not only did it make Carnegie the wealthiest individual in the world, his eye for talent created 43 millionaires among those he appointed to senior positions. Significantly Carnegie looked for evidence of those who had gone the extra mile rather than ask who would go the extra mile.

The starting place in the predictive game was traits. Trait theory had fallen out of favour with most leadership researchers several decades ago who were keen to move away from the "great man (woman) theory" of leadership. However it was traits - now often reinvented as competencies to also incorporate behaviours - that provided the dominant model for prediction.

Although practitioners recognised the impact of contingencies and contextual and situational factors in determining leadership success, faced with the realities of decision making:

- **who do we recruit?**
- **who do we promote?**
- **who do we appoint to senior roles?**

they had to start somewhere. And the somewhere was traits. Tim Judge reviewing the leadership field notes, "as the obituaries (for trait-based theories) were being written, the seeds of a re-emergence were being sown." (8)

What are the attributes and qualities associated with leadership? How are these traits best measured? And how well do they in fact predict future effectiveness?

The first bet in the game: intelligence looks smart - at least to start

The first candidate was intelligence. Despite its promising start in education and then in military selection, intelligence as a construct had became unpopular and its measures had largely fallen into disuse.

It was the pioneering work of Schmidt and Hunter in the 1980s that gave a huge boost to general cognitive aptitude as a potential predictor. In summarising the results from 85 years of employment testing and thousands of research studies, they found that general cognitive ability emerged as a stable and consistent predictor of work performance across a wide range of jobs and occupations, with validity increasing with the complexity of the work task (9).

For a short and inexpensive test to **predict around 25% of effectiveness** for pretty much any role was nothing short of an extraordinary finding.

Given the level of complexity and uncertainty of most leadership roles, it was to be expected that intellectual capacity would also be a good bet as a forecast of leadership effectiveness. However, because the “g factor” of general cognitive aptitude remained too reminiscent of IQ testing and the politics of educational selection and adverse impact, a variety of assessments were constructed as alternatives.

Verbal and numerical critical reasoning tests now dominate the market as a replacement to the “old fashioned” measures of “g” like the Wechsler Adult Intelligence Scale or Raven’s Progressive Matrices. Other test developers took a different route. The Career Path Appreciation methodology was developed as an indicator of capacity to manage the cognitive complexity of progressively more demanding leadership levels. Another research enterprise designed the Cognitive Process Profiler as an attempt to break the “g factor” into multiple cognitive intelligences.

The first bet in the game: intelligence looks smart - at least to start

Two problems faced these test constructors:

The first difficulty for these new assessments was they **still kept measuring general cognitive aptitude** (10) rather than discrete facets of intellectual functioning. While the general factor was expected to break up at higher levels of cognitive capability (11), even within professional and management populations the discrete measures of different intelligences showed substantial inter-correlations.

The label on the tin might suggest a new assessment product. The contents of the tin were still largely made up a massive component of general mental capability. And the "multiple intelligence" research enterprise continues to struggle with the awkward empirical finding that specific cognitive based tests that don't correlate with "g" don't add much to the prediction of leadership effectiveness (12).

The other, more serious, problem was that although cognitive aptitude was proving a decent all-round predictor of work performance, it **didn't fare as well in forecasting differences in leadership effectiveness**. True, intelligence is known to be a "key characteristic in predicting leadership perceptions" (13). But predicting who looks and sounds like a leader is different to forecasting who will be effective as a leader in practice.

When effectiveness was used as the criterion, validity emerged around the .27 mark, i.e. explaining less than 10% of the variance in leadership outcomes (14). A respectable result, if less than sensational, it reminded anyone who knew anything about leadership realities, there was more to leadership effectiveness than the appointment of smart people.

Importantly, intelligence seemed to be a better predictor of leadership effectiveness in low stress situations and where the task required higher levels of direction. Faced with high stress challenges and the need for a more participative style, intelligence seems less critical (or other factors become more critical) to successful leadership outcomes (15).

And the relationship between the intelligence of the leader and that of "followers" appears key; too large a difference between leader and followers and intelligence may be a contra-indicator of effectiveness (16).

Intelligence looks a good bet, most of the time. But the research evidence highlights that we need to think about what specifically we're predicting and the circumstances under which intelligence will be a key component of leadership success, and the situations where it may be less critical.

The second bet: assessment centres flag

Assessment centres, the methodology, originating in officer selection in the military, and then pioneered by AT & T in the 1960s, showed much initial promise (17). Despite the significant cost and time in design and implementation, assessment centres, indicating a predictive power of around 25 - 35%, seemed to go beyond paper and pencil tests of cognitive functioning to broaden the scope of assessment and incorporate additional facets of leadership.

But tracking the meta analysis studies over the last 30 - 40 years, it seems clear that **validity has been steadily falling**. Predictive power had dropped to .37 by the 1980s and the latest review in 2009 estimated validity to be .27 (18).

There had always been a degree of scepticism about some of the claims of the assessment centre method (19). Critics had highlighted the criterion contamination problem in which organisational norms and values influence assessor judgements of candidate suitability, norms and values which also determine who is more likely to get on and progress. Simply put, some "faces fit", both in assessor ratings and in subsequent evaluations of performance and potential.

The critics' arguments were strengthened by the finding that assessment centre validity has always been higher for leadership emergence and progression than against objective measures of leadership impact. Even more troubling was the finding that the predictive power of general cognitive ability may be reduced when used with or alongside assessment centres (20).

The reasons for this decline in assessment centre validity are complex (21), but one factor may be increasing candidate sophistication. In a world of "Assessment Centres for Dummies", YouTube clips and blogs (22) to share not only assessment centre techniques and tips, but the detail of specific exercises used by recruiting organisations, it seems likely that the tactics of the ambitious game-players may be out-manoeuvring their less savvy peers.

Nothing wrong of course with the motivation and wit to prepare for assessment centres, but if it represents a major determinant of assessment centre outcomes, it may go some way to explaining why the take up of assessment centres over the last 20 years or so hasn't seen any major improvement in leadership capability.

If the popular BBC show "The Apprentice" (23) can be seen as an extended quasi-assessment centre (group tasks, presentations, interviews and the like), then the same factors that explain the abysmal hit rate of Lord Sugar's successful candidates over the last six years, probably account for part of the decline in assessment centre validity.

This isn't to argue that we should abandon assessment centre methodology. Assessment centres based on relevant exercises that reflect critical tasks still yield some predictive power. However this flagging validity does require experts and practitioners to revisit their assumptions about design and implementation and re-examine the mechanisms used in data consolidation and selection decision making - and ask why general cognitive ability's power may be reduced - to ensure that the predictive gains still outweigh the high costs of their usage.

The third bet: personality looks encouraging but controversy remains

In 1965, Robert Guion observed: “In view of the problems…one must question the wisdom of using personality as instruments of decision in employment procedures.”

Over 40 years later, personality assessment - and there are now hundreds of options - represents one of the most popular predictors of leadership effectiveness. So what happened? Was there a breakthrough in the research evidence to explain the massive uptake in personality measures? Or have we simply been hoodwinked by smart marketing that exploits our assumptions that personality is in fact more important than it is in reality? (24)

Yes and no.

Yes

A consensus emerged to identify “the big five” of personality - open mindedness, conscientiousness, extraversion, agreeableness and emotional stability - as the building blocks of the structure of personality, a finding which gave impetus to a systematic research enterprise to identify the correlates with work performance. And the pattern looked encouraging; the **Big Five seemed to provide a combined validity of around .39** (25), with open mindedness and conscientiousness (26) highlighted as consistent predictors of task leadership, and emotional stability, extraversion and agreeableness linked to interpersonal effectiveness as a leader.

Significantly, integrity tests seem to have held up well as providing additional predictive power (27), but have proved less popular than personality questionnaires.

No

Critics pointed out that the validity in fact is alarmingly low. Statistical jiggery-pokery can apply any number of adjustments but “there is little controversy that the **uncorrected validities as predictors of performance reported in several existing meta-analyses are close to zero”** (28), and the incremental gains over cognitive aptitude, in fact fairly modest (29).

Specifically the critics argued that there is a world of difference between the proposal that personality shapes leadership behaviour, style and outcomes (where few people would disagree) and the argument that personality profiles from applicants completing self report measures do in fact forecast future leadership outcomes.

Predictive validity for personality questionnaires, the critics argued, cannot be assumed from correlations between personality traits and supervisory ratings of performance within current incumbents.

Here the sceptics suggested two problems:

The third bet: personality looks encouraging but controversy remains

Problem 1: Impression management in selection scenarios

Applicants competing for leadership positions will put their best foot forward to report those qualities they interpret as relevant to leadership success. The fakeability issue has always been the Achilles heel of personality questionnaires. Although some researchers dismiss its impact on validity (30), this seems unlikely (31). There is every suggestion that social desirability is evident among a significant proportion of applicants (32).

Even if social desirability isn't undermining validity, it still represents a problem in selection decision making. In normative-based personnel selection, the applicants who respond with candid honesty are shuffled down the rank ordering in comparison with their more wily and manipulative peers. If this doesn't affect validity, the critics argued, it's still questionable how personality data does in fact predict future performance, at least in any account that is defensible to applicants.

The detailed technical arguments on faking continue.

Do social desirability scales work? Probably not (33).

Do forced choice measures control impression management? Possibly (34), but ipsative measures incorporate additional problems. Smart people seem to be also smarter at improving the scores they identify as job relevant in a forced choice format (35) and direct comparisons across candidates are problematic in ipsative measurement (36).

A review of the faking research concludes somewhat bleakly: "given the complexity of faking in selection settings, it should not be surprising that there are currently few universal recommendations for best practice." (37)

Problem 2: The constraints of supervisory ratings

Supervisory ratings of performance - the common criterion of effectiveness - are not necessarily the best indicator of leadership impact on business performance. We know that line management ratings correlate only modestly with peer and team member evaluations of effectiveness and contribution (38).

We can dismiss this lack of perceptual consistency to argue that peers and team members lack insight into leadership impact. If we do we have to question the fundamental concept of leadership. If team members don't report followership, it's difficult to see evidence of leadership in practice, whatever line management ratings might suggest. In any event there remains a question mark over what supervisory ratings actually capture about leadership effectiveness.

The sceptics argue that supervisory ratings are the result of judgements of who is seen to possess the "right leadership stuff", a set of constructs, based on conscientiousness, agreeableness and extraversion. But "perceived influence is not equivalent to effectiveness, and showing there is a correlation of a personality dimension with perceived influence does not provide a strong basis for use of this measure to select managers who will be effective" (39). Any correlations therefore between personality traits and supervisory ratings, the argument runs, are largely the outcome of a "halo effect" based on line management views who seems confident, hard working, responsive and easy to get on with.

Phil Rosenzweig (40) points out in "The Halo Effect", we get ourselves into a tangle when we allow incorrect attributions of current success to be the predictors of future success.

The third bet: personality looks encouraging but controversy remains

Whatever the lexicon of the different available personality measures (15FQ, 16PF, Dimensions, DISC, Insights, NEO, OPQ, PAPI, Quintax, RPQ, etc) we know from the pattern of inter-correlations that most popular personality tests are variations of the Big Five, albeit to different standards of measurement rigour and efficiency.

At first sight the analysis for personality tests as a predictor of leadership effectiveness looks pretty gloomy. Most fair minded reviewers of the current research database would conclude that personality tests are at best a modest predictor of leadership effectiveness -around 5 -10% - with useful applications for development but incorporating inherent problems in selection situations. At worst, personality testing plays a cynical game, in which integrity and character loses out to guile. Put simply, honest applicants are penalised by peers who deploy smart impression management tactics.

But anyone engaged in the realities of leadership recruitment, career development or succession planning, and alert to the ways in which personality is played out in leadership success, derailment and failure, would be puzzled. Accepting the hazards of impression management and social desirability in selection scenarios, most practitioners would still be baffled by the low power of personality measures to predict leadership outcomes.

For our part we would ask: “what did the researchers expect? “It may be more productive for future prediction to recognise that:

- **the traits of temperament are not the dynamics of personality.** The Big Five and their variants are not the be-all and end-all of personality; instead they map fundamental differences in temperament. As Annie Murphy Paul observes in “The Cult of Personality Testing”: “the Big Five makes me think of being in a spaceship, looking down at the planet below and seeing five continents. That’s useful to know, but once you’re back on earth it won’t help you find your way home.” We need to draw on more sophisticated perspectives (41) to inform our understanding of the factors through which individuals emerge as leaders, make a consistent impact, sustain leadership success over time or derail.

- **all sorts of permutations of personality can achieve leadership success.** “Profiles of Genius”, the detailed examination of outstandingly successful business leaders, highlights the variations in personality profile, everything from the introverted workaholism of Bill Gates of Microsoft to the sociopathic eccentricity of Arthur Jones of Nautilus. It’s difficult to see how any Big Five style personality assessment would have predicted the exceptional success of these very different leaders.

- **different personality themes will be associated with different leadership outcomes.** Summarising leadership effectiveness into a global leadership index is problematic, obscuring the reality that the leadership task varies across different strategic challenges and organisational cultures.

Personality as a bet within the leadership prediction game will no doubt continue. But if it is to become a serious player it will need a rethink, not simply in the design of content but to map more closely how personality dynamics shape effectiveness across the different tasks and challenges of leadership life.

The fourth bet: the rise and stall of emotional intelligence

As the psychometricians puzzled over the evidence, and why personality measures may not have lived up to their original billing, Daniel Goleman (42) joined the game of leadership prediction.

If cognitive and personality measures were doing no more than forecast around 20% of future effectiveness, what accounted for the remaining 80%? All other bets seemed to be off as the forecasting action lay with Emotional Intelligence (EI). EI it's worth noting wasn't introduced simply as a tool to support personal and leadership development. The explicit aim on day one of the enterprise was prediction: EI is "increasingly applied in choosing who will be hired and who will not, who will be let go and who retained, who passed over and who promoted."

Since the mid '90s there has been a proliferation of instruments, and a resulting confusion about the construct, and how it is best defined and measured. For the original pioneers, Salovey & Mayer (43), EI is exactly that: a set of intelligences in the emotional domain to be measured in tests of proficiency; the MSCEIT. With the emergence of competing frameworks and tools from different developers, a flurry of instruments emerged as trait or competency based assessments (e.g. ECI, Bar-On EQI, SUEIT, EIS, IE) measuring everything from optimism and happiness, problem solving and organisational awareness to achievement drive and leading others.

When EI spans such a broad spectrum of attributes, it's difficult to know what it is measuring. Indeed, the low levels of agreement between the different measures of self-reported EI and performance assessments (44) suggests emotional intelligence has become a "catch all" term for a variety of different constructs, with each test developer pursuing a different definition of what EI is and isn't.

It seems that developers followed the advice of pioneer Goleman in his claim that "IQ contributes about 20% to the factors that determine life success, which leaves about 80% to other forces." EI became the "fill in" factor to explain the missing 80% and account for any traits and qualities that could be associated with success.

The first challenge the self report measures of EI had to overcome was **the paradox of the Dunning-Kruger effect**: "that people of low competency in a given area tend to overestimate their abilities, while those of greater competency tend to underestimate their performance."

"If 80% of people believe they are among the top 50% most emotionally intelligent people" (45), and a key component of EI is self awareness, how do those of low emotional intelligence come to the conclusion that they are not emotionally intelligent? For measures of EI as an aptitude this wasn't a problem. For self report competency based EI questionnaires, this issue was never resolved satisfactorily. If normative-based personality measures seem likely to "punish" the honest respondent, then EI self report assessments penalise the self aware.

Undeterred by this measurement problem, the EI enterprise pressed on.

As always, the proof of the pudding is in the empirical eating. With over 15 years of research to review, how does EI fare in the prediction game? After considerable heat between the energetic advocates and the sceptical opponents, some light is emerging.

The fourth bet: the rise and stall of emotional intelligence

Some of the initial findings on first sight seemed positive and impressive, indicating that EI was indeed tapping into an important dynamic of organisational life. EI seemed predictive of work performance, especially where a high level of interpersonal skill was required or in highly stressful jobs.

Next on the list was leadership. After all, leaders with the intelligence to read emotional signals, manage and express their own emotions and connect emotionally to their colleagues, could be expected to be more effective. Indeed if leaders paid more attention to the emotional agenda, the work place would become more productive, rewarding and happy.

15 years on, and we have hundreds of studies, several meta analyses to consolidate the evidence, and scholarly summaries of the findings. Although specific studies based on small samples can be cherry-picked and publicised, the overall pattern of results is dismal. It isn't simply the low level of predictive power that is surprising. Perhaps what astonishes most practitioners is the gulf between the rhetoric of the original claims and the reality of the evidence.

Initial studies claiming sensational findings in fact proved nothing of the sort.

The much acclaimed Van Rooy & Viswesvaran study (46) in reality found that validity was modest for measures of EI - around 0.14 - against objective metrics. And in a kind of head to head competition between IQ and EI, emotional intelligence accounted for only 2% of the variance in effectiveness, whereas cognitive intelligence added 31%.

O'Boyle's meta analysis (47) examining three streams of EI research (EI as aptitude, as self or peer report measures, and as competency based), seemed to find decent levels of validity (.24 to .30). But again, when the data was evaluated to pinpoint the specific incremental validity over and above cognitive aptitude and personality, EI demonstrated next to zero for aptitude measures, 5% for self or peer report measures and only 7% for emotional competency measures.

Maybe the problem lay with the definition of leadership. Perhaps EI wasn't well suited to the nuts and bolts of transactional leadership, and measures of emotional intelligence would do better at predicting the outcomes of transformational leadership.

If in Goleman's view "the major component of leadership is emotional", EI could be expected to do a better job of identifying those leaders better equipped to communicate a vision, generate enthusiasm and trust, and engage others in this mission.

In the Harmes & Crede meta analysis (48), they did find a high level of correlation between self report measures of EI and transformational leadership (i.e. individuals who saw themselves as more effective as transformational leaders also reported themselves as more emotionally intelligent). When the data rather than relying on self esteem, incorporated others' ratings, validity however fell to the .16 mark.

The fourth bet: the rise and stall of emotional intelligence

John Antonakis (49) cuts to the chase in his summary of the evidence, asking:

- does EQ predict leadership effectiveness?
- if it does, does it provide improved prediction over and above established measures?

and concludes: “I have yet to find one study that has followed the accepted guidelines and has shown that EI matters for leadership effectiveness.”

Moshe Zeidner (50) in “What we know about Emotional Intelligence” summarises: “From a practical perspective there is little empirically based evidence, generated from representative samples in different occupational categories, and published in peer reviewed journals to indicate that EI measures do reliably and incrementally predict criteria of job success, beyond that predicted by standard ability and personality measures.”

The conclusion from Christiansen’s 2010 study, looking at EI in a selection scenario: “measures of EI, even performance based measures, offer very little additional information about applicants when measures of cognitive ability and personality are already used in the selection process.” (51)

Another group of researchers have been even more vocal in their commentary of the failings of EI. For them the limited predictive power of EI measures wasn’t just the problem. Their critique focused on the way the emotional intelligence movement attempted to shift the leadership playing field.

When Daniel Goleman said “there is an old fashioned word for EI, it’s character” he was wrong. As Stephane Cote (52) points out “**emotional intelligence is not character.** It’s like any set of skills that we have that can be used to promote moral goals or selfish goals.” The “intelligences” of emotional awareness and self regulation can as easily be deployed by the Machiavellian operator as the authentic leader.

Lynn Waterhouse (53) is even more direct: “nothing in any EI construct precludes someone with high EI from being an immoral person.” Rather than putting character on the leadership agenda, Kristjannson (54) concludes: “EQ lacks moral depth”.

Sir John Bond, former chairman of HSBC warned in 2003: “We’re entering a period where corporate character is going to be ever more important.” As events turned out he was spot on. Unfortunately EI, despite its initial claims to establish character as an important facet of leadership, wasn’t that well positioned to assess it. Even worse, critics propose that the focus on EI may have been a dynamic in taking attention away from the leadership fundamentals of expertise and know-how, and of character and wisdom.

Locke (55), like a cold shower on a winter morning, comments “leadership is not primarily about making people feel good. It’s about knowing what you are doing and knowing what to do.” After a period of “irrational exuberance” maybe we need more leaders who know what they’re doing rather than attempt to deploy emotional intelligence as a replacement for executive intelligence.

The fourth bet: the rise and stall of emotional intelligence

Advocates of emotional intelligence will point out that is too early to draw final conclusions about the predictive power of EI. Here they note that cognitive intelligence has had many decades of debate and dispute (and still does) before the research base established it as a stable and consistent predictor of important life and work outcomes. This of course is fair.

But given the thousands of studies that have now been conducted over a 15 year period, the confusion over the construct, and the well documented psychometric problems of the various assessments, it seems advisable that its initial claims are moderated. And, although many of the early pioneers of the field have been back-pedalling from their original position (56), a quick review of the sites of the distributors of the various commercial EI measures indicates that what empirically is now known about the application of EI to leadership assessment is not being communicated to their users.

The concept of EI did resonate with business leaders looking to engage the work force in challenging times. Arguably, the EI project provided a convenient short-cut that avoided the open discussion about business imperatives, organisational values, leadership wisdom and executive know how. This is the leadership folly of: “we don’t know what we’re doing but we want you to feel positive about whatever we end up doing.”

It’s difficult to conclude anything other than that the EI movement has been a distraction to the enterprise of leadership prediction. “Turf wars over which construct is the true EI” or which measure has or hasn’t been endorsed by Daniel Goleman, exaggerated claims from limited evidence or flawed research have had the impact of heightening cynicism within the leadership population.

The fifth bet: the interview improves or simply makes a good impression

Why can't we simply improve the most popular mechanism we use to select leaders, namely the interview?

Given that no leader will ever be promoted and appointed without some kind of interview, and the reality that selection decision makers will often ignore the "psychometric stuff" anyway if it doesn't confirm their intuitive judgements, shouldn't we concentrate on enhancing the predictive power of the interview process?

As far back as the early 1960s the typical interview was evaluated as an inadequate predictor of work performance (57), providing limited incremental validity to the information (e.g. career history and track record) that was already available to organisations.

No one at this point was prepared to abandon the interview. As Guion and Gibson summarised: "repeatedly discouraging summaries of their reliabilities and validities have not deterred the use of interviews."

Instead a research programme was undertaken to identify the dynamics of the interview process and the range of factors that might constrain its capacity to predict future outcomes. An array of issues were investigated to determine what was holding back reliability and validity, including the social dynamics of first impressions and confirmatory bias, interview-applicant similarity and the impact of non-verbal behaviour, the cognitive distortions of contrast and recency effects, and the influence of different questioning styles.

What did emerge from this research was the **importance of structure** to improve the efficacy of the interview. The attention to greater structure gave rise to two main methodologies:

- competency based strategies focus on the meticulous review of past experience and achievements and the trait/behaviours that underpin success and failure
- situational interviews provide candidates with scenarios relevant to the role to gauge how candidates will tackle the challenge

The emphasis was upon standardisation of interview format and content to provide greater consistency in the evaluation of candidates.

The fifth bet: the interview improves or simply makes a good impression

There is little doubt that the discipline of structure has introduced rigour into the interview process and the evidence indicates improved levels of reliability and validity (58), potentially **of the order of 0.3 to 0.4**, with the indication that interviews which focus on past behaviour provide greater predictive power in leadership assessment than those based on responses to future situations (59).

But.

The first "but" is that in introducing greater consistency the interview has evolved into something that more resembles an examination. Nothing wrong with examinations, but interviews, particularly face to face, may be an expensive way to capture this kind of information. And interviews which discourage the use of follow up and exploratory questions - because they result in "inconsistency" - undermine the key virtue of the interview: the flexibility to engage with the interviewee as an individual.

The second "but "is that structured interviews are rapidly becoming **a game of "show and tell"** in which set questions can be researched easily on the internet (60) and well rehearsed responses prepared. Like the ambitious candidate preparing for an assessment centre, some applicants know how to play the interview game better than others.

A degree of gamesmanship and an element of social grace is of course helpful in most leadership roles; the question is just how much impact these qualities have on leadership effectiveness, and whether or not interview performance over-states this impact. If Jim Collins' analysis of level 5 leadership (61) is correct, then it's difficult, for example, to see how humility plays out in the game of "show and tell" interviews.

Interviews will continue to be part of the repertoire of leadership predictors. The challenge seems to be maintaining that level of structure to establish consistency whilst allowing sufficient diversity and versatility to prevent what should be a dynamic interaction becoming well-prepared and scripted patterns of interviewee response.

The sixth bet: 360° feedback awaits more evidence

360° feedback, the process in which peer, team member and other stakeholder views of leader effectiveness and impact complement the conventional line management perspective, of course has little to offer in the forecast of leadership success among external applicants. Does it have the potential to improve the hit rate for internal candidates as part of talent management activity?

360° feedback processes, at least those that are well designed and implemented (62), have the virtue of work relevance and credibility. They also have the advantage of providing a direct measure of leadership impact in different contexts and assessing how well individuals manage the dynamics of leading up, down and across. If effectiveness is more than a summary of line management perceptions of contribution, 360º feedback seems well placed to reflect a broader set of expectations about leadership priorities and impact.

Initial evidence was promising. Hunter & Hunter (63) in their overview of assessment prediction found a validity of .36 for peer ratings. McEvoy's 7 year follow up study (64) reported that team member ratings predicted managerial performance better than assessment centres.

Furnham & Stringer (65) noted that self-subordinate agreement is related to leader effectiveness, and there is evidence that peer ratings are indicative of the factors that derail leadership performance (66).

Significantly, others' ratings based on Big Five measures provide better predictors of job performance than self report (67), and although self ratings correlated negatively with assessment centre scores (i.e. those who rated themselves highest performed worst), 360º feedback evaluations from others explained a significant part of assessment centre performance (68).

Notwithstanding these encouraging signs, the validity question for 360° feedback in the last 20 years has been less about: does 360° feedback improve our forecast of future leadership effectiveness, and more about: do 360° feedback processes facilitate the kind of behaviour change that improves effectiveness?

Because 360º feedback was largely introduced as a tool to accelerate personal and professional development, shaped by a misguided philosophy that development is more likely to occur when the results are made available only to the participant, there has been a dearth of any long-term validation programmes (69).

The sixth bet: 360° feedback awaits more evidence

We do however have access to a considerable research base to examine the dynamics of 360° feedback, how it works and the specific circumstances in which it makes a positive (or negative) impact. We know:

- feedback content must be credible, focused and designed around a few powerful questions rather than attempt a comprehensive check-list of everything that might be relevant. Long-winded and bloated systems don't work (70).
- which aspects of leadership can be evaluated reliably by the different feedback groups. Line managers can generally evaluate task delivery well, peers are insightful in identifying proactivity and innovation, and team members seem better placed to judge which individuals build trust and provide support. And 360° feedback systems should reflect the differential world views of the feedback groups in the content they present.
- self vs. other agreement is pretty low, i.e. some individuals under-estimate their impact and others over-state their effectiveness and contribution. And where the gap between self and others' ratings is high, individuals are less effective and less likely to improve their performance (71).
- accountabilities need to be established within the different stakeholder groups (individual, line manager, organisation) for change and improvement to happen. The well intentioned strategy in which results are only made available to participating individuals has failed. David Bracken points out the participant "who is not required to share results and action plans with some significant party is unlikely to demonstrate behaviour change". Marshall Goldsmith (72) notes that the potential of 360º feedback to drive improvements in performance is only realised when participants are willing to share the outcomes with their colleagues.

In principle, the application of 360° feedback data - in capturing line management judgements of task delivery, peer views of initiative and innovation, and team member perceptions of credibility and trust - should be an important predictor of future leadership effectiveness. In reality, despite a few encouraging findings, there isn't yet a consistent evidence base to indicate that this is the case.

Now that 360° feedback processes are shifting from the philosophy of "development only" to become integrated within performance and talent management practices, the next five years should highlight how much power 360º feedback can in fact provide to long-term predictions of leadership effectiveness. For the time being we await the evidence.

The seventh bet: judging the situations where situational judgement tests work

As a variation to assessment centres, situational judgement tests (STJs) re-emerged in the 1990s as a potential rival predictor.

Originating in civil service examinations in the U.S. in the late 19th century, situational judgement tests present candidates with work related situations with a range of possible responses to select.

Used for military selection in World War 2, and then taken up within the corporate sector in the 1950s for the "early identification of management potential", situational judgement tests seemed to have been neglected as a serious player in the predictive game until relatively recently.

The initial evidence looked positive. McDaniel in 2001 (73) summarised the data to report a validity of .34, but also noted substantial variation across different studies. Situational Judgement Tests seemed to work better in some formats and situations than others. Because of these differences in design and response task, McDaniel suggests they "could contribute substantially to a predictor composite or offer near zero incremental validity". More recent reviews (74) indicate **validity of around .26**, but also note that most evidence is based on concurrent studies utilising job incumbents rather than applicants.

Current research activity (75) centres now around the optimal design and deployment of situational judgement tests:

- should the situations within the test reflect knowledge (what is the best response?) as a measure of maximal performance, or tendency (what are you most likely to do?) to evaluate typical performance?
- which aspects of leadership can best be assessed within different STJ formats?
- how fakeable are STJs based on the preference format? And how susceptible generally are STJs to practice and coaching effects?
- do written or video-based STJs provide the best predictors?
- how much incremental predictive power do STJs provide over and above cognitive and personality tests?

The seventh bet: judging the situations where situational judgement tests work

Like other predictor options, situational judgement tests come in a variety of shapes and sizes, with undoubtedly a mix of good, bad and ugly applications.

At best, STJs have much appeal. They represent a direct and defensible way of gauging a range of decision making and interpersonal skills relevant to leadership effectiveness. High face validity also means they are viewed positively by candidates. And they seem to report less adverse impact than some other assessment methods.

Perhaps their greatest strength lies in the flexibility of the format to measure specific and distinctive capabilities (e.g. cross cultural intelligence, managing ethical dilemmas) that have often been neglected in leadership assessment. On the other hand, design and implementation costs can be high for role specific STJs. Here significant incremental validity needs to be established to demonstrate their practical utility in leadership selection.

An outside bet: the experience of bio-data

"Did you ever build a model plane that flew?" Responses to this question out-performed the entire battery of psychometric tests as a predictor of pilot training performance (76). Insightful questions that capture information about past experience and achievements may be powerful forecasts of future success.

Biodata (77) and the underlying principle that "the best predictor of what an individual will do in the future is what they have done in the past" is the methodology that rather than ask: "what would you do?", examines "what have you done?" If personality tests operate around a typical response format of: "do you enjoy parties", biodata asks "how many parties have you attended in the last 12 months?" with objective and verifiable response options.

This isn't the argument that the past is our destiny but to suggest that previous patterns of life and work experience have the potential to provide some predictive power.

Biodata hasn't featured much as an assessment tool for leadership effectiveness, despite the evidence that it can be a decent predictor of work performance, at around the .35 mark (78), and that costs are modest compared to say assessment centres and work samples (79).

Although we are interested in the career history and track record of the aspiring and current leaders we evaluate, the specifics of the breadth and depth of their work and general life experience are rarely systematically factored into our judgements and decisions in recruitment and promotion. This is surprising since experience could be expected to be an important factor in accounting for leadership success and failure and career progression. We also know that in selection interviews experience has the largest influence on candidate evaluations (80).

The early research on experience indicated it was a poor predictor. Fiedler (81) noted a validity of .12, arguing that the "link between leader experience and performance has been shot to hell." More recent analysis has been more positive, suggesting **validity is somewhere about .32** (82).

Evaluating the predictive power of experience based on quantity (e.g. number of years or length of service) was always going to be problematic. The challenge (83) was to identify more sophisticated ways of measuring the impact of experience and explaining why different factors of experience may underpin effectiveness. What seems to matter more than length of experience is the density - the intensity of work experience to make a development impact, and timing - when the experience occurs within career development (84).

There are now encouraging signs to suggest we can improve predictive power by being more systematic in our analysis of the experience base of emerging and current leaders:

- attempts to correlate specific dimensions of experience (85) against effectiveness
- McCall & Lombardo's research enterprise identifying the experience factors that shape leadership development and progression (86)
- the growing recognition that character and wisdom (87), neglected facets of leadership life, develop only in the full flow of life experience

An outside bet: the experience of bio-data

Traits, whether as fundamental attributes of the kind assessed in psychometric tests, or as repositioned as competencies, have had a good run for their money in the prediction game. But for the most part they haven't delivered fully against their initial promise.

Experience, and the emerging enterprise to map out the specific factors that accelerate the acquisition of knowledge and know-how, shape attitudes and expectations, and build resilience and wisdom, represents a potentially productive complementary perspective.

Experience as a predictor also resonates well with the growing research programme that has identified the impact of purposeful practice and feedback in building advanced levels of expert proficiency and mastery (88). Exceptional levels of performance, leadership included, emerge from sustained exposure to key challenges with reflective learning to keep shaping improvements in effectiveness.

Fleishmann (89) argues that "the greatest opportunity for advancing our understanding of human performance in organizations lie in examining relationships between life experiences and subsequent job performance". We haven't yet explored fully this opportunity, but our outside bet for improved prediction is the way in which exposure to experience interacts with personality and motivational dynamics to build leadership capability and character. And the more systematic we are in capturing the "density" of life and work experience and accomplishments, the greater the potential to improve our predictions.

Biodata methodology has its critics. Biodata measures, unlike cognitive tests, are less generalisable, and may not "transport" well from one selection situation to another, and validity may "shrink" over time, requiring periodic review and updating. In fact these criticisms reflect the reality of any assessment method. When we assume one predictor will work in all situations, and sustain over time, we get into trouble.

Biodata measures require judgement and insight (and statistical rigour) to design content and scoring algorithms that are appropriate to the candidate population and the specific leadership task being predicted. But in asking "what have you actually done" rather than "what do you think you can do?" biodata represents a grounded alternative to identify the key themes of experience that underpin leadership effectiveness. And this focus on the outcomes of experience rather than the inputs of traits may make it easier to connect leadership assessment to development and progression.

Leadership prediction: a summary

What do we know?

- **it isn't all doom and gloom**. Given the complexity of leadership and realities of organisational change, arguably the power to integrate different predictors to deliver gains of 30% plus isn't in fact bad. Translated into the practical world of leadership selection and promotion, this level of predictive power has the potential to make a positive business difference.

- **we may not however be exploiting current levels of predictive power**. The massive growth within the leadership prediction industry hasn't resulted in a corresponding leadership renaissance. Potentially robust predictors of leadership effectiveness (e.g. biodata to systematically map experience) have not received anything like the attention of other predictors, notably personality measures of the Big Five and emotional intelligence, that in fact have much weaker validity in real life selection scenarios.

- despite the sensational claims **we haven't made any serious inroads into improving** significantly the predictive power of leadership over the last 20 - 30 years. Indeed, some well established and popular methods, like the assessment centre, are showing signs of fatigue.

- a **reliance on individuals to evaluate their own leadership impact is problematic**. The combination of impression management and self deception makes decisions based on personal assessment of leadership effectiveness hazardous. This doesn't necessarily mean we should abandon any predictor based on self report (e.g. interviews, personality tests), but we have to be shrewd in the questions we ask, how we evaluate the answers, and how we integrate this data with other measures.

- **when a new assessment method is claimed as a breakthrough predictor, it probably isn't**. Preliminary breakthrough findings reported as a keynote conference presentation rarely hold up with more systematic evaluation. This isn't to discourage the diversity of innovation in the design of new tools, but it is an appeal to humility to conduct long-term validation before heralding a breakthrough new leadership predictor.

Where does the future of leadership prediction lie?

Looking long term, the landscape of leadership prediction may be very different to the world of 2011. In 10 years time, sophisticated simulations of virtual leadership tailored to specific candidates and roles, or advances in brain science to identify hard-wired cognitive, emotional and motivational patterns, may lead the way in predictive technology.

Alternatively we may have shifted to organisational structures and operating processes that either fundamentally rethink the leadership role or minimise its need, and we'll be in search of a different set of predictors.

What do we need to do in the meantime?

Leadership prediction: a summary

In the short term, we suggest :

- **start with the outcome and work backwards**. We can draw the target around the arrow after it was shot, and we won't be disappointed. But we won't be any wiser in how to shoot future arrows. When we know the specific target we're aiming at, our accuracy might improve.

- **build a theory to connect outcomes to predictors**. Here we map out the link between what we're measuring and the outcomes that matters. And we become more thoughtful in identifying which predictors will be more or less important across different leadership scenarios.

- **a move to specific measures to predict specific outcomes**. Predictors aligned clearly to the desired outcome will fare better than constructs and measures that require a complex chain of cause and effect to explain their predictive power. A bundle of intelligence, conscientious and emotional stability will highlight those individuals who fare reasonably well in organisational life; it won't predict outstanding levels of leadership success in any specific scenario.

- **integrate assessment data wisely to optimise predictive power**. Collecting assessment data is easy. Integrating that data to inform our judgements and decisions is a more challenging task. We gain predictive power when we recognise the strengths and shortcomings of each assessment method to trade off their pros and cons, and look beyond "scores" to see the individual in the context of their past, present and future.

Start with outcomes

Mark Twain noted: “I can teach anybody how to get what they want out of life. The problem is that I can't find anybody who can tell me what they want.” Do we know what we want?

If we hope for the equivalent of a “g factor” in leadership performance - a kind of “p factor” - we might be disappointed (90). Just as attempts to predict aptitude in overall sales effectiveness have largely been unsuccessful (91), outstanding leadership effectiveness won’t be predicted from any standard formula of X + Y = Z.

In “Oops…we did it again”, Seth Kaplan (92) highlights a fundamental weakness in the predictive game. “We start with a predictor and go off in search of something to predict”. And if we go off on sufficient fishing trips eventually we catch some predictive fish.

Stevan Trooboff (93) points out: “the problem in picking leaders lies as much with the definition of what types of leadership are required as in the process of selection itself.” Instead of beginning with the predictor input, it may be more productive to “work backwards” and start with the outcome that is important and valued to the organisation. If we don’t know which leadership outcomes are important or how they impact on organisational performance, it’s difficult to even start the prediction game.

We may want a “black box” assessment that summarises predictive data into a convenient global leadership index. If we do, we may be stuck with modest predictive power. Malcolm Gladwell notes: “We’re used to dealing with prediction problems by going back and looking for better predictors”. Maybe we should rethink our fundamental strategy.

It is only when we start to be more precise about the range of possible leadership outcomes that we can expect to improve our accuracy to predict who will excel.

Which leadership outcomes are important to organisational success?

- short-term productivity or long-term sustainability?
- overcoming the adversity of business turn-around vs. advancing into business expansion?
- specific contribution within a defined role vs. versatility to operate across a spectrum of tasks?
- increasing organisational efficiency or optimising market-place impact?
- employee engagement or broader stakeholder management?
- implementing radical change vs. consolidating incremental improvement?

Ideally we want leaders with the range of experience, skills and motivation to achieve all these outcomes. If leadership is a golf bag (94), super-leaders have the full complement of clubs. Like Jack Welch who went from the persona of “Neutron Jack” in his initial turn-around and down sizing phase at General Electric to business revolutionary, versatile leaders have the adaptability to shift priorities and operating style easily according to the demands of the business situation.

We know there are few super-leaders (95) and assessment processes that assume an abundance of super-leaders will be flawed. Instead, if we’re clear on the outcomes that matter most in a selection scenario, we’ll get better at predicting excellence rather than selecting leaders who just about manage to cope.

Build a theory to link outcomes to predictors

It's unreasonable to expect any assessment method to predict all-singing-all-dancing effectiveness. Leadership comes in too many varieties for this to be meaningful. Instead leadership outcomes vary across different tasks and situational dimensions, and the predictor-outcome linkage will be moderated by:

- organisational **structure** and the degree to which the leadership role is narrowly or broadly defined
- a **culture** that reflects hierarchical status or is more informal and egalitarian
- an **operating ethos** that values experimentation and risk taking or is more conservative and cautious
- **work climate** and the extent to which it is competitive or collaborative
- **followers** are low or high performing employees
- **performance management** systems and the mechanisms that reward short-term results or also recognise long-term contribution

When we start to factor these dynamics into our "predictive formula" we go beyond the simple cause and effect of traits = effectiveness. If we don't address the contextual factors of performance, we will struggle to make sense of why some leaders make an exceptional impact, others despite their initial promise fall by the wayside, and others succeed, despite all objective evidence to the contrary.

When we recognise the interaction of inputs and moderating factors and how they are played out within different organisational challenges, we will get better at identifying who will excel in which leadership situation against which specific outcome.

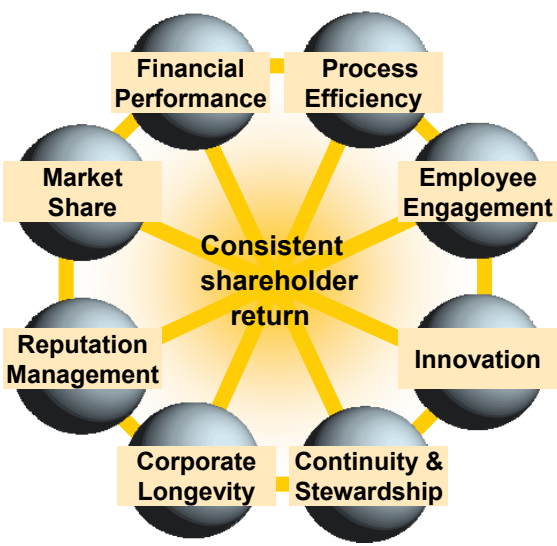
For some leadership scenarios, the predictive game is relatively easy.

When, for example, the leadership challenge is to turn-around a failing business, a successful outcome can be described easily. Here, because the moderating factors comprise a coherent dynamic and the parameters of the task defined, the leadership requirement (knowledge, skills, operating style, etc) can be mapped easily. It's not of course a simple leadership task but prediction is relatively straightforward; potential candidates can be assessed against a clear blue print of success criteria.

For other leadership roles (e.g. acquiring a family run business as part of an ambitious expansion strategy), where stakeholder expectations of success compete, and there is a complex and inconsistent configuration of moderating factors, any number of leadership patterns may or may not prove to be effective. Here the predictive challenge becomes tougher.

Working backwards to build a theory to link outcomes to predictors

Leadership Outcomes



Moderators

Structure
Scope of the leadership role

Organisational culture
Leadership within hierarchical status or more informal and egalitarian

Operating ethos
Risk avoidance vs. experimentation and freedom of discretion

Work climate
Leadership within a competitive or collaborative environment

Follower performance levels
Low vs. high performing team members

Performance management systems
Short-term financial recognition vs. reward for long-term contribution

Challenges (96)

Credibility:
How important is it to establish leadership authority? How easy or difficult will this be?
Which factors will reinforce legitimacy in this leadership situation?

Capability:
Will leadership success hinge on the deployment of a broad range of skills or a focus on a few key strengths?
How important is professional expertise and technical know how vs. generic leadership processes?

Character
Will high levels of integrity help or hinder within this leadership situation?
Does the leadership task require demanding levels of leadership resilience?
What level of challenge vs. conformity is encouraged or discouraged?

Career Management
At what point in the career is this leadership task being faced?
What tactics of self management and political influence will be needed to survive and thrive?

Leadership Inputs

Experience
Exposure to a wide ranging set of life, organisational and business experience

Traits
Fundamental patterns of cognitive and personality that shape capability and preference

Tactics
The deployment of self management stratagems to optimise personal effectiveness and impact within organisational life

The use of specific measures to predict specific outcomes

Morgeson (97) in his critique of the current evidence base of personality measures advocates “the development of custom-developed tests specifically for the job and organisation in question.”

It’s a good general principle. Just as assessment centres or 360º feedback systems are tailored increasingly to reflect the specific leadership themes of most relevance to an organisation, this move implies the abandonment of “off the shelf” products to build targeted tools that reflect the leadership distinctives that matter to different organisations and accommodate the moderating factors of context.

It’s not a new strategy. Almost 50 years ago, Guion & Gottier (98) proposed this strategy of bespoke customised development. Until relatively recently, this was either a high risk or an expensive option. Customised assessments could either be developed quickly and easily resulting in crude measures, or involve lengthy and costly design and construction procedures.

A combination of access to comprehensive item banks with known measurement properties and the application of flexible technology makes this a realistic option for organisations to close the predictor-outcome gap.

Rather than fall back on “off the shelf” assessment centre exercises, situational judgement tests, standard personality tests and the like, this strategy targets assessment content to the specifics of the leadership task being predicted.

The major test publishers and distributors of course dislike this strategy, protesting that bespoke systems of this kind lack demonstrated reliability.

But we’ve tried the off-the-shelf solution for several decades now and claims of robust reliability haven’t done all that much for gains in predictive power. Psychometric elegance might play well in publishers’ manuals or test reviews; it hasn’t improved the accuracy of our leadership forecasts. Does the fact that Test X was developed forty years ago and has a normative base of 400,000 really help the decisions we make today to predict tomorrow’s leaders? (99)

The issue is a shift in thinking to make predictive validity the ultimate criterion of assessment design and implementation. When we move from generic solutions to develop tailored predictor options that are closely aligned to the leadership outcomes that matter to our organisation we make progress. This strategy of bespoke design will also facilitate the development of assessment methods more responsive to cultural diversity across different countries and cultures.

This recommendation of course isn’t the encouragement of more shoddily designed assessment protocols, exercises and tests. It’s an appeal for test developers and practitioners to combine content and technology to incorporate greater versatility and relevance in the measures that are implemented in leadership selection.

Using the available data to improve decision making

Paul Barrett (100) points out: “it’s the actual demonstrated predictive accuracy using real data that calls the shots”. Or put another way, a predictor is only valid in the specific decision scenario in which it is being used. Knowing the validity coefficient of any predictor option may be reassuring but it doesn’t help that much in the task of deciding which candidate to recruit, promote or select for a specific leadership position.

This is the challenge for the practitioner to evaluate the available information, accepting the uniqueness of each candidate and their circumstances, and weight the data to arrive at a judgement that is both insightful and defensible and optimises the odds of future success.

Valid assessments won’t improve the accuracy of leadership promotions and appointments if we:

- **ignore the data**. This is the strategy of impression management to indicate to candidates that rigorous systems are in place to evaluate suitability. The reality is that our subjective intuition (and bias) is, and always was, going to be the key factor in decision making.

- **interpret the data in an ad hoc way**, typically reviewing the available assessment data in a flexible way to make some kind of sense of the results. Rarely does assessment information from different sources (e.g. interviews, assessment centres, 360 feedback, personality measures) line up in a perfectly consistent profile. Invariably there are gaps, ambiguities and contradictions. In ad hoc analysis we end up focusing on the confirming positive data for the candidates we like and highlighting the negatives for those we dislike.

- **rely on our preferred bet**. This interpretative strategy does attempt some kind of weighting of the different streams of assessment data, but does it badly. For example, practitioners in selection and assessment place more reliance on personality tests in hiring judgements than cognitive tests, despite the weaker predictive power of personality based assessments (101).

Using the available data to improve decision making

In high volume recruitment for specific roles with a validated success profile we can design smart algorithms to integrate and weight the available predictive data into clear recommendations from initial screening to first round interviewing and final decision making. This is prediction as a game of odds, utilising the assessment data to improve our hit rate to select candidates with a higher probability of success (and reject those with a higher probability of failure).

Although we can sometimes apply the same logic to entry level leaders (e.g. graduate recruitment programmes) this strategy is unrealistic for most of the leadership selection decisions we face.

The first issue is to work out the **relative power of the assessment methods.**

Multiple assessment methods should of course provide a more valid evaluation than a reliance on any one approach. But only when each method provides distinctive predictive information. Two assessment methods of modest validity, which don't overlap, will out-perform two superior methods which do correlate highly together. Adding a test of EQ to a big five personality questionnaire will do little to improve the accuracy of our predictions. A biodata inventory of experience alongside a test of cognitive aptitude will provide incremental prediction.

The second challenge is the **decision making algorithms we deploy**. Do we apply cut offs for each assessment method to reject candidates below a certain minimum score? Or do we select top down to identify those with the highest scores? Should we base our selection decisions on multiple hurdles in which candidates have to achieve a defined score across the full sequence of assessments? Or do we utilise a compensatory model in which a high score on one assessment can compensate for a lower score on another?

There are different decision making tactics, and the choice will depend on the known validity of each assessment method, the number of candidates and the selection ratio, the organisational impact of success and failure, and importantly, the leadership outcomes that are valued most. Do we, for example, need leadership excellence in a specific area or are we selecting for versatility to operate effectively across a number of different areas?

Here decision making is more the art of professional judgement than the application of actuarial science. The key point is whatever method is utilised to combine and integrate the data from different sources into a final decision, it is established in advance. When selection decision making lacks a clear set of guiding principles, subjective and ad hoc interpretation will wipe out any potential gains from the use of assessment measures, however valid.

Conclusions

Malcolm Gladwell (102) argues; “there are certain jobs where almost nothing you can learn about candidates before they start predicts how well they’ll do before they’re hired.”

Here the argument is that faced with uncertainty and complexity, we just don’t know what is critical to the role or how task requirements might change. The only response then is to allow individuals the opportunity to perform in the role, and have the systems in place to remove quickly those who don’t perform. Easier said then done!

But the point is valid: if we can’t pinpoint the success factors of the role, we should simply flip a coin and save ourselves lots of time and effort.

Where Gladwell (103) is more insightful is the suggestion that we may be asking the wrong questions and looking in the wrong places in our attempts to understand the dynamics of success.

The leadership predictive game, in starting with traits, began in the wrong place.

Rather than taking the time to ask: what outcomes are we attempting to predict, it assumed the goal was a forecast of overall leadership effectiveness, and that the measurement of the right bundle of traits could predict Martini leaders who would be successful “anytime, anyplace, anywhere”.

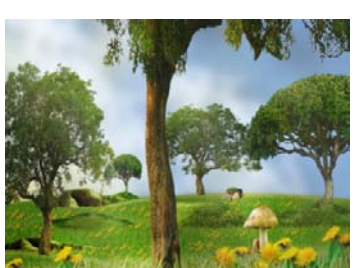
This strategy has proven misguided. Leadership talent doesn’t seem as transferable as first anticipated (104).

If we **begin with leadership outcomes** and the extent to which contextual factors shape performance, we’ll be clearer on the specific factors that predict different forms of success.

And instead of asking: which package of traits provide the “right stuff” to predict effective leadership? (a question that hasn’t advanced predictive power all that much despite over 30 years of sustained research effort) we ask:

- which experiences will shape the kind of character that our future leaders will need to operate with integrity and wisdom?
- what know how and expertise is needed to master the complexities of leadership life?
- which attitudes and expectations are indicative of a leadership outlook that is curious to learn and adapts quickly to new challenges?

we will improve the decisions we make in leadership selection and progression today, and raise the effectiveness of tomorrow’s leaders.



About us

We work with a broad portfolio of clients in the design and implementation of on line services in recruitment and selection; management assessment, development and career management; on line leadership tool kits, 360° feedback, performance management; and talent profiling and succession management.

With over 15 years experience of designing bespoke assessment, development and survey tools for a variety of organisations as well as other consultancy practices, we have also conducted over 200 validation studies, including the use of biodata to predict sales performance in Personal Financial Advisors, the impact of career tactics on progression, 360 feedback data to evaluate cultural competence and its interaction with leadership wisdom, reducing turnover in front line retail employees, and the motivational dynamics of strategic leadership.

If you are interested in our operating philosophy, consultancy services and products for practical assessment and development, do contact us:

- call: 44 (0) 1608 654007
- email: admin@amazureconsulting.com

Or visit our website:
www.amazureconsulting.com

References

1. AM Azure Consulting (2009), "Acceleration: 7 principles to speed up leadership development"
http://www.amazureconsulting.com/files/1/31482697/AcceleratedLeadershipDevelopment.pdf
2. "Only four out of 10 people are satisfied with the quality of leaders in their organisations." OPP survey
3. "Selection & recruitment", Assessment4Potential, 2009
http://www.slideshare.net/LynnTulip/selection-recruitment
4. GoogleAd May 2011 for Hay Group
5. "The measurement outcome equivalence of the Career Path Appreciation"
http://upetd.up.ac.za/thesis/available/etd-03162005-151333/unrestricted/00dissertation.pdf
6. McDaniel M A et al (2005), "Do employment test publisher manuals provide accurate information on test validity", AOM Conference paper
7. The habits of the extra mile
http://www.agiftofinspiration.com.au/stories/achievement/Habits.shtml
8. Judge T (2009), "The bright and dark sides of leader traits", Leadership Quarterly, 20 See also Hoffman B et al (2011), "Great man or great myth? A quantitative review of the relationship between individual differences and leader effectiveness", Journal of Occupational and Organizational Psychology, 84

Intelligence

9. Hunter JE & Hunter RF (1984), "Validity and utility of alternative predictors of job performance", Psychological Bulletin, 96; Hunter JE & Schmidt RF (1998), "The validity and utility of selection methods", Psychological Bulletin, 124; Bertua C et al (2005) "The predictive validity of cognitive ability tests: a UK meta-analysis", Journal of Occupational and Organizational Psychology, 76
10. Ree MJ (1994), "Predicting job performance: not much more than g", Journal of Applied Psychology, 79
11. Brand, C (1997) "The g Factor"
12. This in fact shouldn't have come as too much of a surprise. Guildford's undertaking in the 1950s to split "g" into 150 multiple intelligences had been unsuccessful. Either the specific tests loaded onto "g" or failed to measure much of any predictive value
13. Lord RG et al (1986), "A meta analysis of the relation between personality traits and leadership perceptions", Journal of Applied Psychology, 71
14. Judge T et al (2004), "Intelligence and Leadership: A Quantitative Review and Test of Theoretical Propositions", Journal of Applied Psychology, Vol 89
15. Judge T et al (2004), "The forgotten ones: a re-examination of consideration, initiating structure and leadership effectiveness", Journal of Applied Psychology, 89
16. Bass B (1990), "Bass and Stogdill's Handbook of Leadership"

Assessment Centres

17. Byham W C (1970) "Assessment centres for spotting future managers", Harvard Business Review, 48
18. Gaugler B et al (1987), "Meta analysis of Assessment Centre Validity", Journal of Applied Psychology, 72; Thornton G & Gibbons A (2009), "Validity of assessment centres for personnel selection", Human Resource Management Review, 19
19. Klimoski R & Brickner M (1987) "Why do assessment centres work? The puzzle of assessment centre validity", Personnel Psychology, 40
20. Hermelin E et al (2007), "The validity of assessment centres for the prediction of supervisory performance ratings: a meta analysis", International Journal of Selection and Assessment
21. AM Azure Consulting (2009), "Why assessment centre validity is falling"
http://www.amazureconsulting.com/files/1/29141665/AssessmentCentresValidityAndOptions.pdf
22. Example assessment centre blog
http://www.talkingblues.co.uk/blog/police-assessments/what-the-police-think-about-police-recruitment-assessment-centre-training/
23. The Apprentice and assessment centres
http://www.guardian.co.uk/money/2006/apr/22/graduates.graduation

Personality Testing

24. Gladwell, M (2004, "Personality Plus", New Yorker
http://www.Gladwell.com/2004/2004_09_20_a_personality.html
25. Judge T et al (2002) "Personality and leadership: a qualitative and quantitative review", Journal of Applied Psychology, 87; Ployhart R et al (2001) "Exploring relations between typical and maximum performance ratings and the five factor model of personality", Personnel Psychology, 84
26. Although recently there is the indication that publication bias may be over-estimating the predictive power of conscientiousness. Kepes S et al "Big Five personality and publication bias: conscientiousness worse than assumed" SIOP Paper, 2011
http://www.people.vcu.edu/~mamcdani/Publications/Kepes,%20McDaniel%20&%20Banks%20(2011,%20April).pdf
27. Ones D et al (1993) "Comprehensive meta-analysis of integrity test validities", Journal of Applied Psychology 78
28. Murphy K & Dzieweczynski J (2005), "Why don't measures of broad dimensions of personality perform better as predictors of job performance", Human Performance, 18. See also Morgeson et al (2007) "Reconsidering the use of personality tests in personnel selection contexts", Personnel Psychology, 60; Ones D et al (2007), "In support of personality assessment in organisational settings", Personnel Psychology, 60; Morgeson F et al (2007), "Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection" Personnel Psychology, 60
29. Schmidt F et al (2008) "Increased accuracy for range restriction corrections: implications for the role of personality and general mental ability in job and training performance", Personnel Psychology, 61

References

30. Ones D & Viswesvarn C (1998), “The effects of social desirability and faking on personality and integrity assessment for personnel selection”, Human Performance, 18
31. Schmitt M & Ryan A (1992) “Test taking dispositions: a missing link”, Journal of Applied Psychology, 71; Douglas E et al (1996), “The validity of non cognitive measures decays when applicants fake”, Academy of Management Conference
32. Rosse J (1998), “The impact of response distortion on pre-employment personality testing and hiring decisions”, Journal of Applied Psychology, 83; Weiner J & Gibson N (2000), “Practical effects of faking on job applicant test scores”, SIOP; Birkeland S et al (2006), “A meta-analytical investigation of job applicant faking on personality measures”, International Journal of Selection and Assessment, 14
33. McGrath RE et al (2010) “Evidence for response bias as a source of error variance in applied assessment” Psychological Bulletin, 136; Uziel, L (2010) “Rethinking social desirability scales: From impression management to interpersonally oriented self-control”, Perspectives on Psychological Science, 5
34. Martin B et al (2002) “How effective are people at faking on occupational personality questionnaires”, Personality & Individual Differences, 32; Converse P et al (2008), “Comparing personality test formats and warnings: effects on criterion-related validity and test taking reactions”, International Journal of Selection and Assessment, 16
35. Christiansen N et al (2005) “Reconsidering forced choice item formats for applicant personality assessment”, Human Performance, 18
36. The ipsative debate: issues and objections continue. See, for example, Drpaulatopra, (2010), “Ipsative Tests, Psychometric Properties” http://oprablog.wordpress.com/2010/10/27/ipsative-tests-psychometric-properties/
37. Fluckinger C et al (2008) “Review of faking in personnel selection” In Mandal, M. (Ed). “ In search of the right personnel” http://www.people.vcu.edu/~mamcdani/Publications/Fluckinger,%20McDaniel%20&%20Whetzel%20(2008).pdf
38. Conway J & Huffcutt A (1997), “Psychometric properties of multi-source performance ratings”, Human Performance, 10
39. Morgeson F et al (2007), “Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection“ Personnel Psychology, 60
40. Rosenzweig P (2007), “The Halo Effect”
41. This is more than breaking the Big 5 into sub-scales. It is looking at alternative ways to conceptualise how the dynamics of personality rather than the traits of temperament might impact on work outcomes. See also Rothstein M & Goffin R (2006), “The use of personality measures in personnel selection”, Human Resource Management Review, 16

Emotional Intelligence

42. Goleman D (1995) “Emotional Intelligence”
43. Salovey, P & Mayer, J (1990) “Emotional intelligence”, Imagination, Cognition, and Personality, 9
44. Goldenberg I et al (2006) “The assessment of emotional intelligence: a comparison of performance based and self report methodologies”, Journal of Personality Assessment, 86
45. Brackett M et al (2006) “Relating Emotional Abilities to Social Functioning”, Journal of Personality and Social Psychology, 91
46. Van Rooy D & Viswesvaran C (2004) “Emotional Intelligence: a meta-analytic investigation of predictive validity and nomological net”, Journal of Vocational Behaviour
47. O’Boyle, E. et al (2010), “The relationship between emotional intelligence and job performance: A meta-analysis” Journal of Organizational Behavior, 10
48. Harmes PD & Crede M (2010) "Emotional Intelligence and Transformational and Transactional Leadership - A Meta-Analysis”, Journal of Leadership and Organizational Studies (17)
49. Antonakis, J et al (2009) “Does leadership need emotional intelligence?”, The Leadership Quarterly, 20
50. Zeidner M at al (2009), “What we know about Emotional Intelligence”
51. Christiansen N et al (2010) “Emotional intelligence in selection contexts”, International Journal of Selection and Assessment, 18
52. Cote S (2011), “Dr. Jekyll and Mr. Hyde: the two sides of emotional intelligence” http://www.rotman.utoronto.ca/news/detail.asp?ID=593
53. Waterhouse L (2006) “Inadequate evidence for multiple intelligences, the Mozart effect and emotional intelligence theories” Educational Psychologist, 41
54. Kristjannson K (2006) “Emotional Intelligence in the Classroom”, Educational Theory
55. Locke E (2005), “Why emotional intelligence is an invalid concept”, Journal of Organizational Behavior, 26
56. Daniel Goleman responds to Po Bronson at Newsweek (2009); http://danielgoleman.info/2009/11/03/daniel-goleman-responds-to-po-bronson-at-newsweek/

Interviews

57. Mayfield E (1964), “The selection interview - a revaluation of published research”, Personnel Psychology, 17
58. McDaniel M (1994) “The validity of employment interviews: a comprehensive review and meta-analysis”, Journal of Applied Psychology, 79; Wiesner W & Cronshaw S (1988), “A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview”, Journal of Occupational Psychology, 61; Buckley R & Russell C (1999), “Meta analytic estimates of criterion related validity” in Eder et al “The Employment Interview: Theory, Research and Practice”. But see Schmidt F & Zimmerman FD (2004), “A counterintuitive hypothesis about employment interview validity”, Journal of Applied Psychology, 89
59. Krajewski H et al (2006), “Comparing the validity of structured interviews for managerial level employees: should we look to the past or focus on the future”, Journal of Occupational & Organizational Psychology, 79
60. Interview questions and answers http://jobsearch.about.com/od/interviewquestionsanswers/a/interviewquest.htm
61. Collins J (2005) “Level 5 Leadership: the triumph of humility and fierce resolve” http://hbr.org/2005/07/level-5-leadership/ar/1

References

360 Feedback

62. AM Azure Consulting (2010), "360 Feedback: a shift in agenda" http://www.amazureconsulting.com/files/1/83822555/360Feedback-ANewAgenda.pdf
63. Hunter JE & Hunter RF (1984), "Validity and utility of alternative predictors of job performance", Psychological Bulletin
64. McEvoy G M & Beatty R W (1989), "Assessment centres and subordinate appraisals of managers: a seven year examination of predictive validity", Personnel Psychology, 1989, 42, 68-74
65. Furnham A & Stringfield P (1994), "Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation", Journal of Occupational & Organizational Psychology, 67
66. Gentry et al (2010), " Self-observer rating discrepancies of managers in Asia: a study of derailment characteristics", International Journal of Selection & Assessment, 18
67. Oh In Sue et al (2010) "Validity of observer ratings of the five factor model of personality traits", Journal of Applied Psychology
68. Atkins P & Wood R (2002) "Self versus others' ratings as predictors of assessment centre ratings", Personnel Psychology, 55
69. For example, despite the use of one of our 360 feedback products in a top management programme over the last five years, a dataset of over 4,000 responses, the rules of engagement prohibit any analysis against performance and progression metrics.
70. Kaiser R & Devries K (2003) "Bad items, bad data"
71. Fleenor J et al (2010), "Self other agreement in leadership: a review", The Leadership Quarterly, 21
72. Goldsmith M "Leadership is a contact sport" http://www.marshallgoldsmithlibrary.com/docs/articles/LeaderContactSport.pdf

Situational Judgement Tests

73. McDaniel M & Nguyen N (2001), "Situational judgement tests: a review of practice and constructs assessed", International Journal of Selection and Assessment, 9
74. Lievens F et al (2007), "Situational Judgement Tests: a review of the recent research", Personnel Review, 37; McDaniel M et al (2007), "Situational judgement tests: response instructions and validity: a meta analysis", Personnel Psychology, 60; Whetzel D & McDaniel M (2009) "Situational judgement tests: an overview of current research", Human Resource Management Review, 19
75. Christian M et al (2010), "Situational judgement tests: constructs assessed and a meta analysis of their criterion related validity", Personnel Psychology, 63

Biodata

76. Henry E (1966), "Conference on the use of biographical data in psychology", American Psychologist, 21
77. See Stokes G et al (1994), "Biodata Handbook"; Furnham A & Gunter B (2001) "Assessing Business Potential: a Biodata Approach" as an accessible overview
78. Hunter JE & Hunter RF (1984), "Validity and utility of alternative predictors of job performance", Psychological Bulletin, 96
79. Hunter J (1986), "Cognitive ability, cognitive aptitudes, job knowledge and job performance", Journal of Vocational Behaviour, 29
80. Hitt M & Barr S (1989), "Managerial Selection Decisions Models", Journal of Applied Psychology, 74
81. Fiedler F (1970), "Leadership experience and leader performance", Organizational Behavior and Human Decision Processes, 5
82. McDaniel A et al (1988) "Job experience correlates of job performance", Journal of Applied Psychology, 73
83. Quionones M et al (1995) "The relationship between work experience and job performance", Personnel Psychology, 48
84. Tesluk P & Jacobs (1998) "Toward an integrated model of work performance", Personnel Psychology, 51
85. McCauley, C et al (1994) "Assessing the developmental components of managerial jobs", Journal of Applied Psychology, 79; DeRue S & Wellman N (2009), "Developing leaders via experience", Journal of Applied Psychology, 94
86. "Management Development Through Job Experiences: Bibliography" http://www.ccl.org/leadership/pdf/research/MgmtDevelopmentBib.pdf
87. AM Azure Consulting (2008), "The Seven Pillars of Wisdom" http://www.amazureconsulting.com/files/1/91276339/SevenPillarsofWisdom.pdf
88. Ericson K (2006) "The Cambridge Handbook of Expertise and Expert Performance"; Day, D (2010) "The difficulties of learning from experience and the need for deliberate practice" Industrial and Organizational Psychology, 3
89. Fleishman, E. & Reilly, M (1992). "Handbook of human abilities"

References

The Future of Prediction
90. Schmitt, N et al (2003), "Personnel Selection and Employee Performance", Handbook of Psychology
91. For example: Pendharkar A & Pandey V (2011)"Personality and sales performance" http://www.centraltest.com/ct_fr/upload/documents/SalesProfile_ValidityStudy.pdf
92. Kaplan S et al (2010) "Oops…. We Did It Again", Industrial and Organizational Psychology, 3
93.Troboboff S "Why can't we figure out how to select leaders?" http://hbswk.hbs.edu/item/6103.html
94. Antonakis J (2006) "Leadership: what it is and how it is implicated in strategic change", International Journal of Management Cases
95. AM Azure Consulting (2009), "All in the strategic mind" http://www.amazureconsulting.com/files/1/87912254/AllInTheStrategicMind.pdf
96. AM Azure Consulting (2008) "Rethinking Leadership Realities" http://www.amazureconsulting.com/files/1/73424767/RethinkingLeadershipRealities.pdf
97. Morgeson F et al (2007), "Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection" Personnel Psychology, 60
98. Guion R & Gottier R (1965), "Validity of personality measures in personnel selection", Personnel Psychology, 18
99. Imagine a parallel scenario in which a GP prescribes a drug introduced in 1980. You ask why this specific drug, and she responds: it's a drug I'm familiar with and it's been prescribed now to millions of patients. You ask: but does it work, and the GP answers: not really. In fact there are better treatments but I haven't had the time to be trained in their usage.
100. Barrett P, commentary in "Assessments: can't live with 'em, can't live without 'em" http://www.ere.net/2010/03/31/assessments-can%E2%80%99t-live-with-em-can%E2%80%99t-live-without-em/
101. Christiansen N et al (2010), "Effects of social desirability scores on hiring judgments", Journal of Personnel Psychology, 9
102. Gladwell M (2008) "How do we hire when we can't tell who's right for the job?" New Yorker http://www.newyorker.com/reporting/2008/12/15/081215fa_fact_gladwell?currentPage=all
103. Gladwell M (2008) "Outliers: the Story of Success"
104. Groysberg B (2010), "Chasing Stars: The Myth of Talent and the Portability of Performance" for an analysis of the transferability of talent.