



WHITE PAPER

THE CASE FOR PSYCHOMETRICS IN COACHING

Why coaches, and the organisations that hire them, should put validated measurement at the centre of every engagement.

NSC Group Research · 2026 · approx. 18 min read

THE CASE FOR PSYCHOMETRICS IN COACHING

Why validated measurement belongs at the centre of every engagement

EXECUTIVE SUMMARY

Coaching has scaled into a multi-billion-dollar global industry, yet a stubborn share of engagements still rely almost entirely on conversation, intuition and the coach's personal lens. A growing body of evidence — from International Coaching Federation studies, meta-analyses of executive coaching, and decades of applied psychometric research — shows that coaches and organisations who anchor their work in validated psychometric measurement consistently produce better outcomes: faster behavioural change, higher self-awareness, more accurate development plans, stronger return on investment, and demonstrably reduced risk of harm.

This paper sets out the scientific, commercial and ethical case for psychometrics in coaching, reviews the evidence base, addresses the most common objections, and offers a practical framework for integrating measurement into every stage of a coaching engagement — from contracting through to outcome evaluation.

Headline finding. Across the studies reviewed, coaching engagements that incorporated validated psychometric assessment showed substantially larger effects on goal attainment, self-awareness and leadership behaviour than coaching delivered without measurement — and produced returns on investment typically in the range of five to seven times the cost of the programme (PwC/ICF, 2009; De Meuse, Dai & Lee, 2009; Theeboom, Beersma & van Vianen, 2014; Jones, Woods & Guillaume, 2016).



$5\times-7\times$

reported ROI range on measurement-led coaching engagements

$g = .43-.74$

meta-analytic effects of organisational coaching across outcomes

$r = .14-.35$

typical self-other agreement on leadership competencies

6-12

sessions in the measure-coach-re-measure loop we recommend

WHAT THIS PAPER ARGUES

Coaching without measurement asks the coachee to be both the subject and the only instrument. That is an unusual standard to accept in any other professional discipline. Validated psychometrics introduce a second, independent data source that coach and coachee can both interrogate — and that a sponsor can hold the engagement to at close. The instruments exist and the evidence exists. What remains is the discipline to use them.

Author. Patrick Sheehan, NSC Group. This paper synthesises published peer-reviewed research and industry studies. Where the evidence is contested, we say so.

1. THE COACHING PROBLEM WE DON'T LIKE TO TALK ABOUT

Coaching is built on dialogue. That is both its great strength and its hidden weakness. Without an external reference point, even the most experienced coach is operating inside a shared bubble with the coachee — one in which the coachee's self-perception is the primary data source, and the coach's pattern recognition is the primary instrument. Three well-documented psychological phenomena make this risky.

1.1 The self–other rating gap

Self-ratings on leadership and interpersonal behaviour correlate only modestly with the ratings given by peers, direct reports and managers (Atwater & Yammarino, 1992; Fleenor, Smither, Atwater, Braddy & Sturm, 2010). Meta-analytic estimates of self–other agreement on leadership competencies typically sit between $r = .14$ and $r = .35$. In plain terms: the version of a leader the coach hears in the room is, on average, only weakly related to the version their organisation experiences.

1.2 The Dunning–Kruger effect

People with the lowest competence in a domain tend to overestimate their ability most severely (Kruger & Dunning, 1999). In coaching, this means the coachees who most need development on a given dimension are often the ones least able to nominate it as a goal.

1.3 Coach bias and the ‘favourite tool’ trap

Coaches, like clinicians, are vulnerable to confirmation bias, anchoring, and the well-documented tendency to over-apply whichever framework they trained in most recently (Kahneman, 2011; Passmore & Fillery-Travis, 2011). Without external measurement, the coach has no discipline forcing them to test their early hypotheses.

Validated psychometrics do not eliminate these problems — but they introduce a second, independent data source that the coach and coachee can both interrogate. That alone changes the conversation.

2. WHAT COUNTS AS A ‘PSYCHOMETRIC’?

A psychometric instrument is a standardised, scored measure of a psychological construct — personality, ability, motivation, values, emotional intelligence, behavioural style, engagement or perceived culture. To be useful in coaching, an instrument should meet the standards set out by the International Test Commission (2014), the British Psychological Society (2019) and the American Psychological Association

Reliability. Internal consistency (Cronbach’s $\alpha \geq .70$ as a working minimum, $\geq .80$ preferred for individual decisions) and test–retest stability appropriate to the construct.

Validity. Evidence that the instrument measures what it claims to measure — construct, criterion and content validity — with predictive validity against meaningful outcomes.

Norming. Relevant, recent and representative comparison samples.

Fairness. Evidence of measurement invariance across gender, age and cultural groups, and no adverse impact in selection contexts.

Practitioner competence. Administered and interpreted by a person trained to the appropriate qualification level.

Instruments that fail these tests — including many free online ‘personality quizzes’ and several long-popular type indicators whose psychometric properties have been repeatedly questioned (Pittenger, 2005; Boyle, 1995) — should not be used as the foundation of a coaching engagement, however engaging they feel in the room.

3. THE EVIDENCE: WHAT PSYCHOMETRICS ADD

3.1 Larger and more durable behaviour change

Theeboom, Beersma and van Vianen’s (2014) meta-analysis of 18 organisational coaching studies found a positive overall effect of coaching on performance, well-being, coping, work attitudes and goal-directed self-regulation ($g = .43$ to $.74$ across outcomes). Jones, Woods and Guillaume’s (2016) meta-analysis of 17 workplace coaching studies reported an overall effect of $d = .36$ on organisational outcomes, with notably larger effects where coaching incorporated multi-source feedback — a finding echoed in Smither and colleagues’ (2003) longitudinal work showing that coaching following 360° feedback produced significantly greater improvement in subsequent manager ratings than feedback alone.

3.2 Higher self-awareness — the master variable

Self-awareness is the single strongest leadership development predictor in the empirical literature (Eurich, 2018; Showry & Manasa, 2014). Psychometric instruments raise self-awareness directly by surfacing the gap between how coachees see themselves and how others see them, or between their espoused values and their behavioural defaults. A validated profile gives the coachee something they cannot easily rationalise away — a report with their name on it, scored against a normative sample of thousands.

“Without measurement, the coachee tells you who they think they are. With measurement, you can have a conversation about who they actually are — and the gap between the two is where the coaching value lives.”

3.3 Better contracting and goal setting

Locke and Latham's (2002) goal-setting theory – among the most replicated findings in organisational psychology – shows that specific, measurable, challenging goals produce significantly better performance than 'do your best' goals. Psychometrics give coach and coachee the raw material to set goals with that level of precision: not 'be a better leader' but 'close the 1.8-point gap between my self-rating and my direct reports' rating on psychological safety, and lift my resonance-with-others facet from the 32nd to the 60th percentile over the next two quarters'.

3.4 Predictive validity beyond intuition

Schmidt and Hunter's (1998) landmark 85-year synthesis of selection methods, updated by Sackett and colleagues (2022), shows that cognitive ability tests and well-constructed personality measures consistently outperform unstructured interviews and reference checks in predicting job performance and leadership emergence. The same instruments, deployed in a developmental rather than a selection frame, predict who will benefit most from which kind of coaching intervention (Bono, Purvanova, Towler & Peterson, 2009).

3.5 Demonstrable return on investment

The PwC/ICF (2009) global coaching study reported a median return on investment of seven times the cost of the coaching engagement, with 86 per cent of organisations reporting that they had at least made their investment back. McGovern and colleagues' (2001) earlier MetrixGlobal study put the average ROI of executive coaching at 5.7 times. The common factor in the highest-ROI engagements across both studies was the use of structured measurement at baseline and re-measurement at programme close. Without it, ROI is asserted; with it, ROI is evidenced.

3.6 Reduced risk of harm

Coaching is not risk-free. Studies of 'coaching gone wrong' identify mis-diagnosis of the coachee's issue, coach over-reach into quasi-therapeutic territory, and the imposition of the coach's own model on the coachee as the most common causes of negative outcomes (Berglas, 2002; Schermuly, 2014).

Psychometric assessment helps mitigate all three: it flags when a referral to a clinical professional is appropriate, it anchors the engagement to a shared external reference, and it gives the coachee a basis on which to push back on the coach's interpretation.

4. WHY ORGANISATIONS SHOULD INSIST ON MEASUREMENT

From the buying organisation's point of view, the argument is even sharper. Human resources, learning and development, and talent leaders are increasingly asked to justify development spend against the same standards of evidence as any other operating investment. Measurement-led coaching gives them four things that talk-only coaching cannot.

4.1 An auditable baseline

A validated assessment at the start of an engagement creates a defensible starting point that survives changes in coach, sponsor or reporting line. Without it, the organisation is buying a relationship; with it, the organisation is buying a measurable change programme.

4.2 Portfolio-level insight

When a cohort of leaders is assessed on the same instruments, the organisation can see patterns invisible at the individual level – systemic gaps in psychological safety, consistent over-indexing on task focus at the expense of people focus, divergence between how senior leaders rate themselves and how their workforce rates them.

4.3 Fair comparison across providers

When every coach on an organisation's panel uses the same baseline instruments, the organisation can compare the trajectories of their coachees on a like-for-like basis, rather than relying on the testimonial-style end-of-engagement summaries that have historically dominated coaching evaluation.

4.4 Defensibility under scrutiny

When a board, regulator or shareholder asks what the organisation got for its executive coaching spend last year, the organisation that re-measured at close has an answer. The organisation that did not, has an anecdote.

5. WHAT PSYCHOMETRICS ADD AT EACH STAGE OF AN ENGAGEMENT

5.1 Contracting

Psychometric data sharpens the three-way conversation between coach, coachee and sponsor. It moves the contracting discussion from 'what do you want to work on?' to 'here is what the data is telling us, here is what the sponsor is asking for, and here is what you most want to change – where do those three overlap?'

5.2 Diagnosis

Personality measures (for example Five-Factor Model instruments – Costa & McCrae, 1992), values inventories, emotional intelligence measures (for example the MSCEIT – Mayer, Salovey, Caruso & Sitarenios, 2003), 180° and 360° feedback, and culture and climate diagnostics each illuminate a different facet of the coachee's situation. Used together, they form a multi-method picture far richer than any one instrument, or any conversation, can provide.

5.3 Goal setting

Scores become the anchor for specific, measurable, time-bound goals – exactly the kind Locke and Latham (2002) show produce the largest performance gains.

5.4 Intervention

Knowing a coachee's personality profile, motivational drivers and cognitive style allows the coach to tailor the intervention rather than apply a default model. Coaching is not one-size-fits-all; the evidence that matching intervention style to coachee characteristics improves outcomes is robust (Bozer & Jones, 2018).

5.5 Review and re-measurement

Re-administering key instruments at the close of an engagement – and ideally six to twelve months later – provides the evidence base for ROI claims, identifies whether change has stuck, and signals when a follow-up engagement is warranted.

6. ADDRESSING THE COMMON OBJECTIONS

'Psychometrics put people in boxes.' Good instruments do the opposite. Trait-based personality measures, the Five-Factor Model in particular, describe people on continuous dimensions rather than categories. The 'boxing' critique is largely a hangover from type-based tools that the science has long since moved past (Pittenger, 2005).

'They reduce the human to a number.' The number is the starting point of the conversation, not the end of it. A coach who treats the report as the coachee has misunderstood the tool. The point of the score is to make the conversation more honest, not to replace it.

'They're culturally biased.' Some are. The defence is rigorous selection: insist on instruments with published evidence of measurement invariance across the groups you serve, recent local norms, and a translation and back-translation history. The answer to bad measurement is better measurement, not no measurement.

'The coachee will game the answers.' Well-constructed instruments include validity scales, ipsative elements or

forced-choice formats that make systematic gaming difficult, and in a developmental rather than a selection context the coachee's incentive to game is dramatically lower. In our experience the bigger risk is over-honest self-criticism, not impression management.

'It's too expensive.' A reputable instrument runs to a few hundred dollars per coachee. Set against the cost of a full coaching engagement, and against the five- to sevenfold ROI multipliers reported in the meta-analyses, the cost argument collapses.

7. THE MEASURE-COACH-RE-MEASURE LOOP

We recommend coaches and buying organisations adopt a simple, defensible loop on every engagement.

Step 1 – Baseline. A core personality measure, one situational or 360° instrument, and one context-specific diagnostic (engagement, culture, psychological safety) administered before the first coaching session.

Step 2 – Shared interpretation. A dedicated session in which coach and coachee walk through the data together, surface reactions, and identify the two or three highest-leverage development areas.

Step 3 – Goal contracting. Specific, measurable goals tied to the data and agreed with the sponsor.

Step 4 – Coaching cycle. Six to twelve sessions, with the data referenced explicitly at agreed checkpoints.

Step 5 – Re-measurement. Re-administer at least the situational or 360° instrument and the context-specific diagnostic at programme close.

Step 6 – Sustainment check. Re-measure again at six to twelve months to confirm durability.

8. HOW NSC GROUP APPLIES THIS

NSC Group's coaching practice is built around this loop. Coaching is delivered online across Australia and New Zealand, and every engagement begins with measurement rather than with a conversation about what the coachee would like to work on.

Where an engagement calls for multi-rater, normed instruments – personality, emotional intelligence, leader interpersonal practice, psychological safety and speak-up, or high performance culture – we draw on the validated diagnostic suite published by Influential (influential.solutions), whose instruments are engineered specifically for the measure-coach-re-measure cycle described in this paper. Reports are benchmarked, cohorts are comparable, and every re-measurement is traceable to its baseline.

The result, for the sponsor, is an engagement that can be evaluated on evidence rather than on impressions — and for the coachee, a second mirror that makes the first one honest.

9. CONCLUSION

The evidence is no longer ambiguous. Coaching engagements that integrate validated psychometric measurement produce larger behaviour change, higher self-awareness, sharper goal contracting, better-matched interventions, demonstrable return on investment and lower risk of harm than coaching delivered on conversation alone. For the coach, measurement is a discipline that protects the coachee from the coach's own blind spots. For the buying organisation, measurement is the difference between purchasing a relationship and purchasing a change programme.

The case for psychometrics in coaching is, in the end, the case for taking coaching seriously as a professional practice rather than a craft. The instruments exist. The evidence exists. What remains is the discipline to use them.

TALK TO NSC GROUP

Measurement-led coaching, delivered online

NSC Group provides executive and leadership coaching, delivered online across Australia and New Zealand, with validated measurement at baseline and at close. nscgroup.au/coaching · info@nscgroup.au · +61 401 625 132

REFERENCES

All sources below were consulted in preparing this paper and are publicly locatable. Where the evidence is contested, the text says so at the point of citation.

1. AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.
2. Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45(1), 141–164.
3. Berglas, S. (2002). The very real dangers of executive coaching. *Harvard Business Review*, 80(6), 86–92.
4. Bono, J. E., Purvanova, R. K., Towler, A. J., & Peterson, D. B. (2009). A survey of executive coaching practices. *Personnel Psychology*, 62(2), 361–404.
5. Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist*, 30(1), 71–74.
6. Bozer, G., & Jones, R. J. (2018). Understanding the factors that determine workplace coaching effectiveness: A systematic literature review. *European Journal of Work and Organizational Psychology*, 27(3), 342–361.
7. British Psychological Society. (2019). Psychological Testing Centre: Test user standards. BPS.
8. Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Psychological Assessment Resources.
9. De Meuse, K. P., Dai, G., & Lee, R. J. (2009). Evaluating the effectiveness of executive coaching: Beyond ROI? *Coaching: An International Journal of Theory, Research and Practice*, 2(2), 117–134.
10. Eurich, T. (2018). What self-awareness really is (and how to cultivate it). *Harvard Business Review*, January.
11. Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly*, 21(6), 1005–1034.
12. International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. ITC.
13. Jones, R. J., Woods, S. A., & Guillaume, Y. R. F. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, 89(2), 249–277.
14. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
15. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
16. Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717.
17. Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, 3(1), 97–105.
18. McGovern, J., Lindemann, M., Vergara, M., Murphy, S., Barker, L., & Warrenfeltz, R. (2001). Maximizing the impact of executive coaching: Behavioral change, organizational outcomes, and return on investment. *The Manchester Review*, 6(1), 1–9.
19. Passmore, J., & Fillery-Travis, A. (2011). A critical review of executive coaching research: A decade of progress and what's to come. *Coaching: An International Journal of Theory, Research and Practice*, 4(2), 70–88.
20. Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research*, 57(3), 210–221.
21. PricewaterhouseCoopers & International Coach Federation. (2009). ICF global coaching client study: Executive summary. ICF.
22. Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection. *Journal of Applied Psychology*, 107(11), 2040–2068.
23. Schermuly, C. C. (2014). Negative effects of coaching for coaches: An explorative study. *International Coaching Psychology Review*, 9(2), 165–180.
24. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, 124(2), 262–274.
25. Showry, M., & Manasa, K. V. L. (2014). Self-awareness — key to effective leadership. *The IUP Journal of Soft Skills*, 8(1), 15–26.
26. Smither, J. W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource

feedback ratings over time? *Personnel Psychology*, 56(1), 23–44.

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

Where validated multi-rater measurement is required, we work alongside Influential Solutions (influential.solutions).

ABOUT NSC GROUP

NSC Group is an Australian workforce culture, organisational review and performance consultancy led by Pat Sheehan. We work with boards, executives and leadership teams on culture diagnostics, organisational review, capability frameworks and executive coaching. Learn more at nscgroup.au.

© 2026 NSC Group. This paper may be shared in full with attribution.

NEXT STEP

Where to take this

If you are commissioning coaching, ask what will be measured before the first session is booked.

TALK TO US ABOUT COACHING

Coaching, delivered online

Coaching is delivered online across Australia and New Zealand, with validated instruments used where the sponsor wants the engagement evaluated on evidence rather than impressions.

nscgroup.au/coaching

THE SERVICE BEHIND THIS PAPER

Executive and leadership coaching

How we scope, run and hand over the work.

nscgroup.au/services/executive-and-leadership-coaching

Or start a conversation

Pat Sheehan, NSC Group · p.sheehan@nscgroup.au · nscgroup.au/contact