

Measuring What Matters

What empathy assessment still isn't capturing — and why it matters for real business change

Stuart Nolan — Stuart Nolan Consulting · 2026
stuartnolan.com · stuart@stuartnolan.com

— A B S T R A C T

Organisations are investing in empathy development as never before — driven by the pressures of AI transformation, remote and hybrid working, and growing evidence that human relationship quality is the primary determinant of organisational performance. But most of that investment is built on a measurement foundation that has a significant structural gap. The dominant tools for assessing empathy — the Interpersonal Reactivity Index, the Empathy Quotient, and the Reading the Mind in the Eyes test — are well-constructed instruments that have advanced our understanding of the cognitive and emotional dimensions of empathy. What they do not assess is the physical dimension: the somatic attunement, involuntary mirroring, and body-level responsiveness that determines whether empathy actually changes how someone behaves in the room with another person. This paper reviews the state of the art in empathy measurement, identifies what remains systematically uncaptured, and argues that this gap has direct and costly consequences for organisations trying to measure, develop, and demonstrate the value of empathy as a business capability. Stuart Nolan Consulting is actively developing new approaches to address this gap. This paper describes the problem those approaches are designed to solve.

We have been measuring empathy for decades. We have been missing a key dimension.

The Measurement Gap	The Consequences	What Is Needed
<p>The most widely used empathy assessment tools — the IRI, EQ, and RMET — measure cognitive awareness and emotional recognition. None of them measure physical empathy: the somatic attunement that determines whether empathic understanding actually changes behaviour.</p> <p>Systematic reviews of over 6,000 studies consistently conclude the same thing: no empathy measure currently available is both psychometrically robust and comprehensive. The field's own researchers describe the absence of a gold standard as a defining problem.</p>	<p>When organisations commission empathy assessments, they receive a partial picture. The dimension most likely to explain why empathy training does or does not produce behaviour change is the one that is not being measured.</p> <p>Training programmes designed around self-report data develop the dimension of empathy that is easiest to measure, not the dimension that is hardest to fake and most directly connected to performance outcomes.</p>	<p>A valid assessment of empathy as an organisational capability requires an instrument that captures all three dimensions of the Threefold Model: cognitive, emotional, and physical. Current tools cover the first two. The third is unmapped.</p> <p>Stuart Nolan Consulting is developing proprietary approaches to address this gap. This paper describes the gap those approaches are designed to fill.</p>

— THE CORE ARGUMENT

The absence of a gold standard for empathy measurement is not a minor technical inconvenience. It means that every empathy training programme built on a self-report baseline is measuring the dimension of empathy that people can most easily misrepresent — and ignoring the dimension that most directly determines whether they behave differently when it matters.

— THIS PAPER COVERS

1. Why measurement matters for organisations

The commercial and strategic stakes of measuring empathy well — or badly.

2. What the dominant tools measure

A fair account of the IRI, EQ, and RMET — what they measure well and where their limits lie.

3. What the field itself says

Systematic reviews, meta-analyses, and the field's own conclusions about the absence of a gold standard.

4. The physical dimension gap

Why the physical dimension of empathy is systematically absent from measurement — and what that costs.

5. What a complete assessment would look like

The dimensions a valid empathy audit needs to cover, and the criteria a good instrument must meet.

6. The consequences for training design

Why measurement drives training design — and what happens when training is designed around incomplete measurement.

7. What Stuart Nolan Consulting is developing

The proprietary measurement approaches currently in development, and how to get involved at an early stage.

8. Practical application

Programme formats, pricing, and how to begin.

Why Measurement Matters for Organisations

Empathy has become a serious organisational priority. The language has shifted: where development programmes once spoke about communication skills or emotional intelligence as peripheral competencies, they now speak about empathy as core infrastructure — for leadership effectiveness, for AI adoption, for employee retention, for the quality of the human relationships that determine whether strategic objectives are actually realised. This shift reflects the evidence. The research on empathy and organisational performance is not marginal. It is robust, replicated, and growing.

But seriousness of intent is not the same as rigour of execution. Organisations that commission empathy assessments, design training programmes, and measure outcomes are only measuring what their instruments are capable of measuring. If those instruments have significant gaps — and the evidence reviewed in this paper demonstrates that they do — then the investment may be considerably less effective than the data suggests. Worse, it may be optimising for the wrong thing entirely.

The stakes are practical. When an organisation uses a pre/post empathy assessment to evaluate a training programme, it is asking: did this intervention produce a measurable change in empathy capability? If the instrument does not measure the dimension of empathy that the training was designed to develop, the answer it returns is meaningless. Not wrong — meaningless. It is measuring something real, just not the right thing.

The three questions measurement should answer

For an empathy assessment to be genuinely useful in an organisational context, it needs to answer three questions. First, what is the current empathy capability baseline across the organisation — and where are the gaps? Second, did the training or intervention produce a real change in that capability? Third, does the change in capability translate into different behaviour in the situations that matter?

Each of these questions makes a different demand on the measurement instrument. The baseline question requires breadth: the instrument needs to cover all the relevant dimensions of empathy, including the ones that are hardest to measure. The outcome question requires sensitivity: the instrument needs to be responsive to genuine change over the timescales of a training programme, not just stable enough to produce consistent scores. The behaviour question requires ecological validity: the instrument needs to measure empathy as it operates in real conditions, not just in the controlled conditions of a questionnaire or laboratory task.

The instruments most commonly used in organisational empathy assessment do not fully meet any of these three demands. This paper explains why — and describes what a better approach would need to include.

6,000+

Studies reviewed in a 2024 PLOS One systematic review

of empathy and compassion measurement instruments

0

Instruments identified as both psychometrically

robust and conceptually comprehensive

45+

Years since the IRI was introduced — still the

most widely used empathy measure in practice

What the Dominant Tools Measure

Three instruments dominate empathy measurement in research and applied settings. Each represents a genuine contribution to the field. Each has real strengths. And each has specific, documented limitations that matter enormously when the instrument is used as the basis for organisational development decisions. A fair account of these tools requires acknowledging both.

The Interpersonal Reactivity Index (Davis, 1980)

The IRI is the most widely used general-purpose empathy measure in existence. Developed by Mark Davis and first published in 1980, it has been used in thousands of studies across psychology, medicine, education, and organisational research. Its longevity is a mark of genuine quality: it is well-constructed, freely available, and offers a multidimensional view of empathy that single-score instruments cannot match.

The IRI measures four subscales: Perspective Taking (the tendency to adopt others' psychological points of view), Empathic Concern (feelings of sympathy and compassion for others), Personal Distress (personal discomfort in response to others' distress), and Fantasy (the tendency to imaginatively transpose oneself into fictional situations). This structure captures something real about the cognitive and emotional dimensions of empathy.

Its limitations are equally real. The IRI is a self-report instrument, which means it measures what people believe about their own empathy, or what they are willing to report about it, not what they actually do. Self-report empathy scales are particularly vulnerable to social desirability effects — empathy is widely understood to be a positive quality, and people reliably present themselves as more empathic than they are. Research published in the Stanford Encyclopedia of Philosophy's entry on measuring empathy notes that self-report scales show no significant correlation with actual empathic accuracy: people who score highly on empathy questionnaires are not reliably better at reading other people. This is not a minor technical qualification. It means the IRI measures dispositional self-assessment, not demonstrated capability.

The IRI also does not assess the physical dimension of empathy at all. Its four subscales address cognitive inference, emotional resonance, and imaginative projection — all processes that occur above the threshold of conscious awareness. The somatic dimension, which operates below that threshold and which neuroscience identifies as the foundation of genuine empathic response, is entirely absent from its design. This was not an oversight by Davis: the physical dimension was not a central concern of empathy research in 1980. But it is a central concern now, and the IRI cannot fill that gap.

The Empathy Quotient (Baron-Cohen & Wheelwright, 2004)

The Empathy Quotient was developed by Simon Baron-Cohen and Sally Wheelwright at the University of Cambridge, originally as a tool for studying empathy differences in adults with

Asperger syndrome and high-functioning autism. It has since become widely used in occupational and clinical settings as a general measure of empathic tendency. The EQ defines empathy as a combination of cognitive drive — the motivation to understand another’s thoughts and feelings — and affective response — the emotional reaction to another’s state. This is a theoretically coherent framework, and the EQ has reasonable internal consistency and discriminant validity for its originally intended purpose.

Its limitations in organisational contexts are threefold. First, like the IRI, it is a self-report instrument subject to all the social desirability limitations that entails. Second, it was designed to identify empathy deficits at the clinical end of the distribution — to distinguish individuals with known empathy impairments from neurotypical adults. Its sensitivity at the higher end of the empathy distribution, where most development work operates, is considerably weaker. A tool designed to identify floors is a poor instrument for measuring ceilings. Third, it shares the IRI’s absence of any physical dimension: the EQ does not measure somatic attunement, postural mirroring, or bodily co-presence. These are not part of Baron-Cohen’s theoretical framework, which focuses on mentalising and affective resonance.

A 2024 PLOS One systematic review of empathy and compassion measures, covering over 6,000 studies, found that instruments including the EQ — while widely used — consistently fail to cover all domains of the empathy construct, and that none currently available can be considered psychometrically robust and comprehensive.

The Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001)

The RMET is a performance-based measure rather than a self-report scale, which represents a genuine methodological advance. Participants are shown photographs of the eye region of faces and asked to select which of four mental-state descriptors best matches what the person is thinking or feeling. Because it tests what people can actually do rather than what they report about themselves, it avoids the social desirability problem that undermines self-report instruments.

The RMET has been cited over 2,000 times and has contributed meaningfully to research on theory of mind and social cognition. For the specific purpose of measuring the ability to recognise emotional states from facial cues, it has reasonable test-retest reliability and some convergent validity with related constructs.

Its limitations, however, are substantial and well documented in recent literature. A 2023 study in a large, demographically representative US sample of 1,181 adults identified serious psychometric problems, including inadequate factor structure, poor internal consistency, and the finding that nearly a quarter of the test items did not meet the criteria for inclusion that were established in the original validation study. A 2024 review surveying 1,461 studies that administered the RMET concluded that the validity evidence was seriously lacking and called for a moratorium on its use in research. The debate continues actively in the

literature, with proponents and critics publishing their assessments and counter-assessments through 2024 and 2025.

More fundamentally for organisational purposes, the RMET measures a specific cognitive skill – the ability to identify mental states from static photographs of isolated eye regions. Even if it measures that skill reliably, that skill is not the same as empathy. Recognising that a photograph of a pair of eyes depicts someone who is ‘contemplative’ rather than ‘preoccupied’ tells us nothing about whether the person can attune to a colleague who is struggling in a meeting, read the physical tension in a room before it becomes conflict, or make someone feel genuinely present with them rather than merely processed. The ecological gap between what the RMET tests and what empathy requires in practice is wide. Baron-Cohen himself acknowledged this limitation in the original 2001 paper: the stimuli are static, whereas real social interaction involves continuous, dynamic movement. The RMET was designed as a research instrument for a specific clinical purpose. Its application as a general organisational empathy measure takes it well beyond its validated scope.

INSTRUMENT	WHAT IT MEASURES	DIMENSION COVERAGE	KEY LIMITATION
Interpersonal Reactivity Index (IRI) Davis, 1980	Self-reported cognitive and emotional empathic tendencies across four subscales	Cognitive ✓ Emotional ✓ Physical ✗	Self-report only; scores do not correlate with actual empathic accuracy; physical dimension entirely absent
Empathy Quotient (EQ) Baron-Cohen & Wheelwright, 2004	Self-reported cognitive drive and affective response to others’ emotional states	Cognitive ✓ Emotional ✓ Physical ✗	Designed for clinical floors, not development ceilings; self-report bias; physical dimension absent
Reading the Mind in the Eyes Test (RMET) Baron-Cohen et al., 2001	Ability to identify mental states from photographs of the eye region	Cognitive (recognition only) ✓ Emotional ✗ Physical ✗	Static images; poor internal consistency; measures recognition, not response; no ecological validity for real interaction
Perth Empathy Scale (PES) Brett et al., 2023	Cognitive and affective empathy across positive and negative emotional valence	Cognitive ✓ Emotional ✓ Physical ✗	Most recent and rigorous self-report scale, but still self-report; physical dimension not addressed

What the Field Itself Says

The limitations described above are not external criticisms of these instruments. They are the field's own findings, documented across systematic reviews, meta-analyses, and the ongoing methodological debate that the empathy measurement literature has been conducting with itself for years. The problem is not that these tools are bad. The problem is that the field has repeatedly identified their gaps, described what a better approach would require, and — to date — not produced it.

The evidence is consistent and accumulating. A systematic review published in *Frontiers in Psychology* (2021), examining the psychometric quality of instruments available to assess empathy across the decade from 2009 to 2019, found that while many tools had undergone rigorous development and testing, none could be recommended as a gold standard. The conclusion is direct: no instrument currently available is both psychometrically robust and comprehensive in its coverage of the empathy construct.

A 2024 *PLOS One* systematic review and evidence map, covering over 6,000 studies and examining 559 papers reporting on the development or validation of empathy and compassion measures, reached the same conclusion. The size of the literature is not matched by its quality: the field produces instruments at a rapid rate, but the instruments consistently fail to cover all relevant dimensions of the construct, and consistently lack the responsiveness to change that organisations need when using them as outcome measures for training interventions.

The responsiveness problem

Responsiveness to change is the psychometric property that matters most in an organisational development context, and it is the one that receives least attention in the empathy measurement literature. An instrument used to evaluate a training programme needs to be sensitive enough to detect genuine change over the programme's duration — typically weeks or months, not years. A 2021 systematic review of nursing empathy measures found that responsiveness was tested in only three of the instruments examined. A workplace training meta-analysis published in *BMC Psychology* (2024) found that when empathy training outcome studies use the IRI or Jefferson Scale of Empathy as their primary measure, effect sizes are modest — but whether this reflects modest training effects or instruments that are insufficiently sensitive to genuine change is not possible to determine from the data.

This matters commercially. An organisation that runs a physical empathy training programme and measures outcomes using the IRI may see limited pre/post change — not because the training did not work, but because the IRI is not measuring the dimension that changed. The training produces change in somatic attunement and physical responsiveness. The IRI measures self-reported cognitive perspective-taking and emotional concern. These are related constructs, but they are not the same, and change in one does not reliably

produce change in the measured score of the other.

The self-report ceiling problem

There is a deeper structural problem with self-report instruments in organisational settings that goes beyond social desirability. Empathy, as understood through the Threefold Model developed from doctoral research at Lancaster University, is partly an automatic, pre-conscious process. The physical attunement that determines whether someone actually feels heard happens below the threshold of deliberate thought — it is a bodily process that precedes cognitive interpretation. Self-report instruments, by definition, can only access what is available to conscious reflection.

This creates an assessment ceiling that cannot be resolved by improving the items. No matter how carefully a self-report scale is constructed, it cannot measure what happens in the body before the mind has time to engage. A manager who has developed strong somatic attunement through physical practice will not necessarily score higher on the IRI than a manager who has done a reflective practice programme — because both are reporting on their conscious self-perception, and the physical change produced by somatic training is not primarily accessible to self-report.

Research on empathic accuracy — the ability to correctly infer what another person is thinking or feeling in real time — confirms this structural limit. Multiple studies have found that empathic accuracy is “not fully captured by self-report measures of empathy,” and that the relationship between self-report empathy scores and actual performance on empathic accuracy tasks is weak or absent. The Stanford Encyclopedia of Philosophy’s review of empathy measurement summarises the finding bluntly: researchers following Ickes speculate that the lack of correlation “indicates that people in general have little meta-knowledge regarding their empathic ability.”

Most existing measures of empathy rely on self-reports of dispositional tendencies or assess subjective or physiological responses to static images; consequently, they fail to assess the ability to monitor rapidly changing social cues, a skill that is very important in navigating real-life social interactions.

— Zaki & Ochsner, 2011 — cited in multiple systematic reviews

The Physical Dimension Gap

The Threefold Model of empathy distinguishes three dimensions: cognitive empathy (the intellectual understanding of another’s perspective), emotional empathy (the felt sharing of another’s affective state), and physical empathy (the somatic attunement through which the body responds to another’s physical presence before conscious interpretation occurs). The first two dimensions are measured, imperfectly but meaningfully, by the instruments described above. The third is not measured at all.

This is not a minor omission. The neuroscience of empathy consistently identifies the physical dimension as foundational: it is the layer that operates first, fastest, and most automatically. The mirror neuron research of Rizzolatti and colleagues, the neural basis of emotional contagion, the documented effects of physical co-presence on emotional attunement — all of this describes a dimension of empathy that is body-based, pre-cognitive, and inaccessible to self-report. It is also the dimension that is most directly trainable through physical practice, most directly relevant to what organisations need empathy to do, and most completely absent from the available measurement toolkit.

Why physical empathy is missing from the measurement literature

The absence is not accidental. Measuring physical empathy is genuinely difficult. It requires instruments that can detect somatic responses, postural mirroring, and attunement dynamics that operate below the threshold of conscious awareness — in real interaction, under real conditions, with all the noise and complexity that real conditions introduce. Laboratory paradigms that strip out that complexity (static photographs, controlled video clips, questionnaire items) cannot capture what they strip out.

The research infrastructure that has grown up around the IRI, EQ, and RMET reflects a methodological convenience as much as a theoretical commitment. These instruments are easy to administer, easy to score, and easy to include in studies with student populations. They have generated a large body of literature because they are tractable research tools. The dimension of empathy they systematically miss is the one that is hardest to measure with a questionnaire and a university psychology pool.

This is the characteristic pattern of measurement-driven science: we understand most thoroughly what we can measure most easily, and we design training programmes to develop what we can demonstrate has changed. If the measure does not capture physical attunement, training programmes are not designed to develop it. If training does not develop it, the capability gap persists. If the capability gap persists, the organisations that most need physical empathy — those navigating AI-driven transformation, managing remote teams, dealing with trust deficits and adoption failures — continue to invest in the dimension of empathy that is least likely to produce the change they need.

What physical empathy measurement would need to assess

A valid instrument for the physical dimension of empathy would need to assess several specific capabilities that current tools do not reach. It would need to measure how accurately a person reads physical and postural cues in real interaction: not whether they can identify a mental state from a static photograph, but whether they can track the continuous, micro-level physical signals through which another person communicates their emotional state below the threshold of deliberate expression.

It would need to measure ideomotor responsiveness: the degree to which a person's body automatically mirrors and resonates with the physical state of another person in their presence. This is a measurable phenomenon — it can be detected through behavioural observation, postural analysis, and physiological correlation — but it requires observation of real interaction rather than administration of a paper instrument.

It would need to be sensitive to the change produced by physical empathy training. The neural circuits underlying somatic attunement are plastic: they can be developed through structured physical practice. An instrument capable of measuring the physical dimension of empathy needs to be sensitive enough to detect that development over the timescale of a training programme — a criterion that no existing instrument meets, because no existing instrument measures this dimension at all.

And it would need to be deployable in practical organisational settings — not just in neuroscience laboratories with fMRI equipment and trained researchers. The instrument needs to be rigorous enough to support ROI measurement and programme evaluation, and practical enough to be used by L&D teams without specialist technical infrastructure.

What a Complete Assessment Would Look Like

The limitations of current measurement tools do not mean that measurement is impossible or that organisations should stop assessing empathy. They mean that a complete empathy audit requires a more sophisticated and multi-method approach than a single questionnaire can provide. The design principles for such an approach are clear from the literature, even if a fully validated instrument does not yet exist.

Dimensions that must be covered

A complete empathy assessment needs to span all three dimensions of the Threefold Model. The cognitive dimension — perspective-taking, the ability to understand another’s point of view — can be partially addressed through performance-based tasks that test accuracy rather than self-report. The emotional dimension — affective resonance, the capacity to feel something in response to another’s emotional state — requires instruments that go beyond self-report toward observable physiological or behavioural markers. The physical dimension — somatic attunement, the body-level responsiveness that precedes conscious interpretation — requires direct behavioural observation in real or realistic interaction conditions.

No single existing instrument covers all three. The most rigorous current approach combines the best available self-report scale — the Perth Empathy Scale (PES, 2023) is currently the most psychometrically defensible option, with its attention to both positive and negative emotional valence and its cleaner separation of cognitive and affective dimensions — with performance-based tasks that test accuracy in realistic conditions. But even this combination leaves the physical dimension unmeasured.

Properties a good instrument must have

Beyond dimensional coverage, a valid empathy assessment instrument needs to meet four psychometric criteria that the literature repeatedly identifies as absent from current tools. First, responsiveness to change: the instrument needs to be sensitive enough to detect genuine development over the timescale of a training programme. Second, ecological validity: what the instrument tests needs to correspond to what empathy looks like in the settings where it matters — management conversations, team dynamics, client relationships, high-pressure decision-making. Third, resistance to social desirability: the instrument needs to produce scores that reflect actual capability rather than self-presentation. Performance-based tasks have a significant advantage over self-report here. Fourth, relevance to business outcomes: the scores need to predict something that organisations care about — adoption rates, retention, engagement, conflict frequency — so that investment in development can be connected to measurable returns.

The Perth Empathy Scale represents the best current attempt to address the first three criteria within the self-report format. It was designed with explicit attention to social

desirability bias, covers both positive and negative emotional valence (a dimension previous scales neglect), and has demonstrated measurement invariance across gender. But it remains a self-report instrument, and a 2023 self-report instrument cannot measure what physical empathy requires.

The case for multi-method assessment

The direction the field is moving — slowly — is toward multi-method assessment that combines self-report data with performance-based and observational methods. This approach is more resource-intensive than a single questionnaire, but it is also more valid. The combination of a strong self-report scale (for cognitive and emotional dimensions), a behavioural accuracy task (for cognitive empathy as demonstrated rather than reported), and an observational protocol for physical attunement (for the physical dimension that self-report cannot reach) would produce a genuinely comprehensive picture of empathy capability.

The organisational value of this is direct. An assessment that produces a dimensional profile — showing where in the Threefold Model an individual's empathy capability is strong and where it is underdeveloped — enables targeted training design. It identifies not just that empathy development is needed, but which dimension needs most work and what kind of intervention is most likely to be effective. This is the diagnostic precision that organisations are currently unable to access, because the tools they use do not distinguish between dimensions.

The Consequences for Training Design

Measurement shapes training. This is not a theoretical observation — it is a practical mechanism. When organisations design empathy development programmes, they design them to produce measurable outcomes. What is measurable determines what is prioritised. What is prioritised determines what gets developed. What gets developed is assessed by the available instruments. If the available instruments do not measure the physical dimension, the physical dimension does not become a programme objective. If it is not a programme objective, it is not trained. If it is not trained, the most foundational and least fakeable dimension of empathy remains undeveloped.

This is the cycle that the current measurement gap perpetuates. It is not a conspiracy and it is not negligence. It is the natural consequence of measurement-driven programme design operating with an incomplete measurement toolkit. The evidence for this pattern is visible in the training literature: a 2024 systematic review and meta-analysis of 50 workplace empathy training interventions found that the overwhelming focus of implemented training methods was on cognitive-verbal techniques — perspective-taking exercises, reflective practice, feedback-based approaches. Physical empathy methods were absent from the reviewed interventions.

What cognitive-only training produces

Cognitive empathy training produces genuine, measurable change in self-reported perspective-taking and empathic concern. The meta-analytic effect sizes are real. Participants who complete well-designed reflective practice programmes do understand others' perspectives more consciously, do articulate empathic concern more fluently, and do score higher on the IRI subscales that measure these dimensions. The training is not ineffective for what it measures.

What it does not reliably produce is change in the physical dimension — the automatic, pre-cognitive attunement that determines whether someone feels genuinely heard rather than intellectually processed. A manager who has completed an empathy awareness programme may score significantly higher on the EQ without having developed any new physical responsiveness. They understand more, they articulate more, they reflect more. But when they walk into a room where a team member is in distress, their body responds the same way it always did. The cognitive layer has been upgraded. The physical layer has not been touched.

This matters because the physical layer is the one that the other person experiences. Cognitive empathy produces better understanding. Physical empathy produces a different quality of presence — the kind that people describe when they say someone really listened to them, or really saw them, or made them feel understood in a way that went beyond what was said. That quality is not produced by perspective-taking exercises. It is produced by the physical development that current training largely ignores, because current measurement

does not direct attention toward it.

The ROI measurement problem

The measurement gap also undermines the ability to demonstrate return on investment for empathy development programmes. If the primary outcome instrument does not capture the dimension of empathy that the training was designed to develop, ROI calculations are built on incomplete data. A programme that produces significant change in physical empathy capability but shows modest change on an IRI pre/post assessment will appear to have limited impact – generating precisely the scepticism that leads organisations to deprioritise investment in the dimension of development most likely to produce durable behaviour change.

This creates a commercial as well as a methodological problem. Organisations that invest in serious empathy development and then measure outcomes with self-report instruments that cannot detect the change produced are likely to undervalue the investment. The people who participated know something changed – they can feel it. But the measurement does not support the financial case for continuing or expanding the programme. Without better measurement, better training cannot fully prove its value.

What Stuart Nolan Consulting Is Developing

The gap described in this paper is not a gap that Stuart Nolan Consulting identified and then searched for a solution. It is a gap that became evident through the training work itself.

Participants in physical empathy programmes consistently report changes that they cannot fully articulate — a shift in how they are with other people, a new quality of attunement, a change in what they notice and how they respond. These are real changes. The neural circuits underlying physical attunement are plastic — they respond to structured practice in the same way that any developed physical skill does.

This creates two problems simultaneously. For participants and the organisations that commissioned their training, it limits the ability to quantify and communicate the value of what has changed. For the field more broadly, it means that the most promising direction in empathy development — the physical dimension — remains uncharted in measurement terms, making it harder to build the evidence base that would drive wider adoption.

Stuart Nolan Consulting is actively developing proprietary approaches to address this gap. The methodological development is proceeding along three connected lines.

A performance-based assessment for physical empathy

The first line of development is a performance-based instrument — distinct from self-report — designed to assess the physical attunement dimension of the Threefold Model. This instrument is being developed around the specific capabilities that physical empathy training produces: the ability to read postural and physical cues in real interaction, ideomotor responsiveness to another's physical state, and the quality of physical co-presence that distinguishes trained from untrained attunement. Because it is performance-based rather than self-report, it sidesteps the social desirability problem entirely. What it measures is what someone can do, not what they say about themselves.

The validation approach uses populations who have and have not completed physical empathy training as the criterion groups — a natural design that is made possible by the established training methodology and the documented evidence of behaviour change that the training produces. This is methodologically clean: the instrument needs to be sensitive to the difference between trained and untrained physical empathy, and the training programme provides the contrast group.

An organisational Empathy Audit

The second line of development is an organisational diagnostic that aggregates individual empathy profiles into a team and organisational capability map. The goal is to give L&D directors and HR leads a dimensional picture of empathy capability across their organisation — not an average score, but a map that shows where in the Threefold Model different parts of the organisation are strong and where development is most needed.

This Empathy Audit is designed to answer the three questions identified in Section 1: it

provides a baseline, enables outcome measurement after interventions, and produces scores that can be connected to organisational performance data. The dimensional structure means it can support targeted programme design rather than generic empathy awareness training. An organisation whose audit reveals strong cognitive empathy but underdeveloped physical empathy needs a different intervention from one whose profile shows the reverse.

Longer-horizon development

A third line of development, at an earlier stage, explores whether patterns in how people interact with AI systems can provide a passive behavioural signal related to empathy capability. The theoretical foundation for this is established: the Media Equation (Reeves & Nass, 1996) demonstrates that people respond to AI and computer systems as social actors, triggering real social behaviours including the kind of attentive, attuned responses that are characteristic of high-empathy individuals. The question being explored is whether variation in those responses across a workforce, aggregated anonymously at team and organisational level, could provide a complementary data source for the Empathy Audit without requiring separate assessment sessions.

This is genuinely novel territory and is described here as a research direction rather than a current product. But it is the kind of methodological innovation that becomes possible only when the theoretical gap has been clearly identified — and the identification of that gap is what this paper is designed to provide.

Founding partnership

The development of validated assessment instruments benefits from diversity of context. Stuart Nolan Consulting is working with a small number of founding partner organisations in the development of the Empathy Audit methodology. Founding partners receive early access to proprietary methodology before it is publicly available, direct influence over its development, and preferred pricing that is locked in permanently. This is not a research arrangement in which organisations contribute data to someone else's project — it is a formative collaboration in which partners help shape something that will become a significant tool in the field. If this is of interest, the conversation begins with a free 30-minute discovery call.

Practical Application

For organisations that recognise the measurement gap described in this paper, there are practical steps available now — before the fully validated new instruments are complete. The most important is to stop using single self-report instruments as the sole basis for empathy programme evaluation. The IRI and EQ are useful as one element of a broader assessment, not as the primary evidence that a programme has worked. Supplementing self-report data with behavioural observation, manager-rated outcome measures, and business performance indicators gives a more complete picture than any single questionnaire can provide.

The second step is to ensure that the training itself addresses the physical dimension, not just the cognitive and emotional ones. If training is designed exclusively around reflective practice, perspective-taking, and emotional awareness work, it will not develop the physical attunement that the measurement gap has made invisible. Stuart Nolan Consulting's programmes are specifically designed to develop the physical dimension — the one that most training ignores, and the one that determines whether empathy changes how someone actually behaves in the room.

FORMAT	WHAT IT IS	BEST FOR	FROM
Keynote Talk	An experiential introduction to physical empathy and the measurement gap, designed for leadership conferences, all-hands events, and L&D forums. 30–90 minutes, any group size. Participants experience the evidence in their own bodies.	Leadership teams, HR and L&D directors, organisations at the start of an empathy development programme.	£1,500
Workshop or Lab	A half-day immersive session developing physical empathy capability through structured practice. Suitable for leadership teams, management cohorts, and cross-functional groups of 8–30 participants.	Teams where the measurement gap is contributing to limited training outcomes, or where behaviour change is needed rather than awareness change.	£3,500

Empathy Audit	A multi-method diagnostic of current empathy capability across the Threefold Model, including the physical dimension. Provides a dimensional profile at individual, team, and organisational level, with bespoke programme design. Currently available as an early-partner engagement.	HR and L&D leads building the business case for empathy investment, organisations whose previous empathy training has not produced the expected behaviour change.	£4,500
Full Programme	A multi-cohort programme integrating all three dimensions of the Threefold Model with measurement checkpoints, leadership coaching, and ROI tracking. 6–12 months.	Enterprise teams, NHS trusts, and organisations with annual transformation budgets seeking durable, measurable development outcomes.	£12,000

A note on measurement in programme design

Stuart Nolan Consulting designs measurement into every programme from the start, not as an afterthought. This means agreeing in advance what change looks like and how it will be evidenced, using a combination of methods appropriate to the programme’s objectives, and connecting individual-level development to the team and organisational performance indicators that justify the investment. As the proprietary assessment instruments described in Section 7 come online, they will be integrated into this measurement framework. Organisations that engage with SNC now are well positioned to adopt these tools as they become available.

Conclusion: Measuring What Empathy Actually Does

The empathy measurement literature has been honest about its own limitations for decades. Systematic reviews consistently find no gold standard. Researchers consistently identify the gap between what instruments measure and what empathy requires. The academic debate about the validity of specific tools, particularly the RMET, is active and unresolved. The field knows the instruments are incomplete. It has continued to use them because the alternatives are harder, slower, and more expensive to develop.

That situation is now commercially costly in a way it was not before. Organisations are investing in empathy development at scale, under real performance pressure, in contexts where the measurement gap translates directly into wasted budget and missed outcomes. A programme that develops cognitive empathy awareness and measures it with a self-report scale may show respectable effect sizes without producing any change in the physical attunement that determines whether a manager actually feels present to their team, whether a change process produces genuine buy-in, or whether AI adoption rates improve because people trust the humans leading the transition.

The physical dimension of empathy is not peripheral. It is foundational. The neuroscience is clear on this. The training evidence is clear on this. What has not been clear — until the development work described in this paper begins to produce results — is how to measure it. That clarity is coming. In the meantime, organisations that want to get ahead of this should look carefully at what their current assessment instruments are actually measuring, and ask honestly whether the training they are commissioning is developing the dimension of empathy that most needs development — or the one that is easiest to measure.

The question is not whether empathy can be measured. The question is whether we are measuring what empathy actually does.

— Stuart Nolan, Stuart Nolan Consulting

— — WORK WITH STUART NOLAN CONSULTING

Empathy training that changes behaviour.

Stuart Nolan Consulting works with senior leadership teams, HR directors, and learning and development functions across the UK and internationally. Every engagement begins with a free 30-minute discovery call to understand your specific context, the empathy gaps most relevant to your organisation, and what a programme designed for your situation would look like.

Keynote Talk · From £1,500 — Experiential introduction for events and leadership conferences.

Workshop or Lab · From £3,500 — Half-day immersive team training, 8–30 participants.

Empathy Audit · From £4,500 — Multi-method diagnostic + bespoke programme design, organisation-wide.

Full Programme · From £12,000 — Multi-cohort, 6–12 months, with measurement and ROI tracking.

Get in touch

stuart@stuartnolan.com · stuartnolan.com

— — ABOUT THE BOOK

How to Train an Empath: Lessons from a Professional Mindreader (Stuart Nolan, Billet Publishing, 2025) sets out the full evidence base, methodology, and practice framework for the physical empathy training described in this series — including complete case studies, participant accounts, and implementation tools. Available at stuartnolan.com/book

— REFERENCES

Sources cited in this paper

- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger Syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163–174.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The ‘Reading the Mind in the Eyes’ test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241–251.
- Brett, J. D., Becerra, R., Maybery, M. T., & Preece, D. A. (2023). The psychometric assessment of empathy: Development and validation of the Perth Empathy Scale. *Assessment*, 30(4), 1156–1169.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126.
- Emslander, V., & Scherer, R. (2024). Categories of training to improve empathy: A systematic review and meta-analysis. *Psychological Bulletin*. doi: 10.1037/bul0000453.
- Higgins, W. C., Ross, R. M., Langdon, R., & Polito, V. (2023). The ‘Reading the Mind in the Eyes’ test shows poor psychometric properties in a large, demographically representative US sample. *Assessment*, 30(6), 1849–1863.
- Ickes, W. (Ed.). (1997). *Empathic accuracy*. Guilford Press.
- Nolan, S. (2025). *How to Train an Empath: Lessons from a Professional Mindreader*. Billet Publishing.
- Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Rodrigues, C., Ferreira, A., & Figueiredo, A. (2021). Empathy: Assessment instruments and psychometric quality — a systematic literature review with a meta-analysis of the past ten years. *Frontiers in Psychology*, 12, 781346.
- Shin, H., Park, J., Lee, E., & Jeong, Y. (2024). Training emotional competencies at the workplace: A systematic review and meta-analysis. *BMC Psychology*, 12, 692.
- Sinclair, S., et al. (2024). Measures of empathy and compassion: A scoping review. *PLOS One*, 19(1), e0297099.