

Evaluating Practitioner-Led Workplace Interventions

Paul Barrett

Chief Research Scientist, Cognadev UK Ltd.
Honorary Professor of Psychology, University of Auckland, NZ
Adjunct Professor of Psychology, University of Canterbury, NZ

Roché Herbst

CEO: Human Capital Solutions Inc., Canada
Registered Organizational Psychologist



NZ Psychological Society Annual Conference, Nelson, 2014

When organizational psychologists formally evaluate the validity of an 'intervention', it is virtually always in the context of assessment e.g. interview, assessment centre, or questionnaire deployment. The primary expression of validity is the Pearson correlation coefficient between assessment variables and specific outcomes or supervisor ratings. But how do we assess the validity of probably the majority of consultant or practitioner-led 'psychologically-oriented' interventions within an organization, such as identifying 'potential', team-building, team-dysfunction interventions, executive coaching, leadership-development, personal skills enhancement, encouraging self-insight, emotional intelligence development etc.? That is the focus of this presentation.

The starting point is the acceptance of a basic premise: **in the workplace, the judgement of the validity or otherwise of an assessment lies with those who initiated and paid for the intervention.**

This is in stark contrast to the typical academic exposition of 'types of validity'. But, we deal directly with this issue at the start of this presentation. Given we accept that premise, validity evaluation (if it is not to be a matter of scientific import), becomes a matter of aligning the expectations of the purchaser of an intervention with observed outcomes.

Our thinking is that validity evaluation for any workplace intervention has to be model-based. That is, we ask **the client** (the purchaser/responsible decision-maker; i.e. the person who will be held accountable for the success or otherwise of the purchased intervention) three questions, build an evaluation model around the answers, and then evaluate expectations against observed outcomes:

- ① What exactly do you expect to happen as a result of deploying the intervention?
- ② Over what time-period do you expect to observe the effects?
- ③ What, for you, will constitute failure?

By "What exactly do you expect to happen as a result of deploying the intervention?" we mean just that. Not broad generic 'vapourware' statements, but clearly elaborated outcomes, even if those outcomes are in fact the result of a hypothesised diffusion process. i.e. nothing specific is expected as the outcome; instead, the outcome will be observed as the result of a diffusion process which has a synergistic impact on many areas in an organization over time.

Validity evaluation requires matching expectations of specified outcomes to actual observed outcomes.

An example of what constitutes an intervention model for a 3-day residential executive leadership course is presented here, along with the schematic principles of deploying this evaluation model in the workplace.

Definitions

The Practitioner: the person selling a service or product to a client within an organization.

The Client: the “responsible” person who buys the time, service, or product from a practitioner.

An Intervention: anything of which advertises it’s beneficial impact on an organization’s human capital (the workforce, potential or otherwise).

2

The Practitioner: This includes I/O psychologists, test-publisher/distributor sales staff, independent HR or management consultants, counsellors and coaches. Indeed, all those who sell psychological services or products into organizations on the basis that what they are selling will solve a problem, create value, enhance employee productivity, or otherwise benefit the organization in some way.

The Client: The person or persons within an organization who make purchase decisions for psychological or indeed, any kind of intervention deployed on the workforce. They are responsible for incurring a cost, and presumably will be required by others, or perhaps by themselves, to justify that cost in terms of demonstrating the benefit or otherwise of what they have purchased. In essence, the client is the arbiter of whether the claims of benefit associated with the intervention, as stated at the point of sale by the practitioner, were valid.

An Intervention: Examples are team-building sessions or away-days/adventure trips, employee coaching and development, assessment strategies such as the use of psychometric tests, the use of assessment centres, enhancement of psychological states such as well-being, leadership courses, enhancement of psychological attributes such as emotional intelligence or employee engagement.

Workplace Utility and Validity

Workplace Utility: the judgement of the validity or otherwise of an assessment lies with those who initiated and paid for the intervention. If the client says “it works for me”, and is able to show others evidence in support of their judgment, then “job done”.

This kind of validity is not to be confused with that associated with the measurability of an attribute.

3

① **Does the assessment actually ‘work’ in practice?** Put another way, do purchasers/users of the assessment find it useful, to the extent that they are willing to pay for it on a continuing basis (*not just a one-shot sale but a commitment to keep purchasing/using it*)? There is a lot of huffing and puffing and all manner of “*you can’t be serious*” arguments thrown at assessments like the DISC, Enneagram, and MBTI. But, clearly many users find them valid, to the extent that they are willing to pay again and again to use them. If these assessments led to adverse workplace consequences, would the sellers of these products be able to remain in business and be as profitable as some are? From this ‘workplace utility’ perspective, the assessments are adjudged valid if they can be shown to possess enduring utility for clients, *as adjudged by the clients*, not by psychometricians or 3rd parties drawing conclusions from abstract perspectives. **The question of course is whether clients simply ‘believe’ they work (as a matter of faith), or make a reasoned decision based upon empirical evidence of ‘effect’.**

② **Measurability.** This scientific (not psychometric) perspective is concerned with answering: “does this assessment ‘measure’ what it is claimed to measure?” This implies that the rules for instantiation for the meaning and variation of the attribute are consistent with the procedures used to construct ‘measures’ of it. Here, whether an assessment is adjudged ‘valid’ or not is determined by the evidence that the psychological attribute proposed as being ‘measurable’ actually *is* measurable in terms of its specified attribute properties (*it varies as a quantity or as a series of orders*). This is where experimentation is required, establishing that what is proposed as causal for variation in the attribute can be observed using methods and procedures derived from the rules for instantiation of the meaning and measurability of the particular attribute. E.g. if I claim ‘*this is Sales Potential and it is measured like this*’, I am required to establish that variation in the attribute is causal for variation in those behaviours which I have claimed are the observable ‘result’ of possessing magnitudes of Sales Potential.

Borsboom, D., Mellenbergh, G.J., & Van Heerden, J. (2004). [The concept of validity](#). *Psychological Review*, 111, 4, 1061-1071.

Borsboom, D., Cramer, A.O.J., Kievit, R.A., Scholten, A.Z., & Franic, S. (2009). [The end of construct validity](#). In Lissitz, R.W. (Eds.). *The Concept of Validity: Revisions, New Directions, and Applications* (Chapter 7, pp. 135-170).

Michell, J. (2009). [Invalidity in Validity](#). In Lissitz, R.W. (Eds.). *The Concept of Validity: Revisions, New Directions, and Applications* (Chapter 6, pp. 111-133). Charlotte: Information Age Publishing.

How is it done? A simple example

A commercial organization sends 15 future leaders per year on a 3-day residential executive leadership course at a cost of ~\$5,000 per person = \$75,000 p.a. cost.

The CEO (prompted by the CFO) asks the HR director:
 “Are you sure this is better than just grooming these people in-house by moving them into leadership roles in a structured employment trajectory? We’ve been doing this for 5 years – and frankly I can’t see any impact/difference from when we started this- except we have spent nearly \$0.5million – and for what?”

4

The question posed by the CEO is likely one posed by many external to the purchaser of such ‘product’, although informally. And to be honest, \$75,000 may not be a significant amount for an organization making profits of say \$500million a year. But in an organization struggling financially and/or commercially, or where such expenditure has to be accounted for in a publicly funded organization, that \$75,000 takes on more ‘significance’.

This is what one NZ seller of a residential 3-day leadership course claims will be the results of employees attending it:

- Understand the difference between leadership and management
- Know the four things most expected of leaders
- Have identified what’s at the source of effective and powerful leadership
- Have far greater self awareness and leadership wisdom
- Have created your leadership backbone – your leadership purpose, vision and values
- Have a comprehensive DISC behavioural profile report showing your primary leadership style and how to adapt your style for best results
- Know how to create shared purpose, vision and goals for your organisation or team
- Know how to lead change
- Know how to create a high-performing team
- Know and have practised fundamental coaching skills
- Know how to delegate effectively and in a way that develops others
- Know the five keys to motivating those you lead
- Have an action plan to transfer learning back to the workplace

To begin the evaluation model design, a consultant now sits down with the client and asks the three big questions.

The three simple questions to be asked of the client

- ① What exactly do you expect to happen as a result of deploying the intervention?
- ② Over what time-period do you expect it to occur?
- ③ What, for you, will constitute failure?

5

What exactly do you expect to happen as a result of deploying the intervention? This from a headline in a recent Workforce Week ..

<http://www.workforce.com/archive/feature/training-development/special-report-leadership-development-corporate-leaders/index.php>

Special Report on Leadership Development: Corporate Leaders Train in Fire Drills and Funny Skills

To sharpen their leadership skills, managers from about two dozen companies in New York and New Jersey joined with some of their employees to play firefighter one afternoon last May. The four-person teams traded in their white collars for gas masks and "turnout gear"—traditional firefighter garb. Under the supervision of New York City firefighters, the teams rushed into burning buildings, rescued passengers from simulated subway accidents or performed other high-pressure emergency drills.

The justification from HR: "Since our training puts people into crisis situations, *hopefully* they will be better prepared to handle any crisis that arises in the workplace." Look at that word "hopefully" in the context of the \$2,500 cost per 4-person team taking part.

This is where the probing by the consultant of the purchaser of the product is both interactive but keenly focused on what the intervention is meant to produce in terms of consequences; observable outcomes which will directly or indirectly affect organizational performance, and not necessarily what the seller has indicated (*e.g. they will get along with one another better, they will now understand themselves and others better, they are more engaged, they understand what makes great leaders etc*).

In addition, some explanatory computational simulations may be required to predict the expected consequences of seller-provided validity coefficients quoted in support of the proposed deployment of assessment psychometrics, or the use of assessment centre tasks which possess ambiguous empirical predictive accuracies/validities.

The goal is always to help inform the client as to the risk and benefits that may accrue from a deployment, and to establish a formal evaluation model if a decision is made to purchase and deploy an intervention.

Q1: What exactly do you expect to happen?

1. Know how to create shared purpose, vision and goals for your organisation or team.

Evaluation: 3rd-party observations/Ratings of Team Performance where the attendee leads a team, compared to observations made prior to course attendance.

2. Know how to delegate effectively and in a way that develops others

Evaluation: 3rd-party observations/Ratings of delegation performance, compared with previous ratings.

6

From the claims made by the leadership course seller, these are the only two which relate directly to performance in the workplace .. The rest are no more than vague *"isn't life wonderful"* statements which are meaningless in the context of answering the simple question: *"what exactly is meant to happen, that can be observed and evaluated in the leadership performance of the attendee in the workplace, as a result of the course?"*

As my alter-ego Sir Alan Sugar might say: *"This is not a game"* . It is about spending money on employees that can be shown to increase the probability of someone evolving into a leader within the organization, as against doing something else, or nothing at all. This course could be a good wheeze sold for a bucketful of money to clueless clients.; then again it might be excellent value for money.

Some might respond: *"No-one really knows what makes a good leader, but surely anything we do to broaden our "potential's" theoretical knowledge, self-insight, motivational skills, and leadership wisdom etc. must be a good thing?"* Perhaps so .. but if you have no evidence that doing any of this does actually 'make a difference', then they are simply taking a punt – this time a \$75,000 p.a. punt. Perhaps because they fear that doing nothing will be far worse than doing something. And, if enough 'significant others' are doing the same in a area, decision-making becomes easy, justified to any 3rd party as *'this is what everyone else is doing, and they can't all be wrong'*. The problem of course is that what may work for one organization may not work for theirs.

Some may reply that organizational schedules and workloads prohibit such evaluation; you just have to trust the large consultancy reviews and those who claim *'doing X is best practice'*, and do it. Here the client is simply evading any judgment of their own by stating: *"I'm taking a punt, because X or Y say it's a good punt"*. Sometimes we all have to take a punt; but in many cases we can evaluate its utility. When it's an expensive punt, and you can evaluate it but ignore the opportunity to do so, is it because intervention-evaluation is seen more as a potential threat than a possible benefit?

Q1: What exactly do you expect to happen? What else?

3. Better ratings of trust and satisfaction with their leadership from subordinates .
4. Greater profitability or lower costs in their business unit.
5. Evidence of greater productivity from those they lead.
6. More innovations emanating from their team.
7. Evidence of more strategic input (*ideas, implementations, new systems*) into the wider organization.
8. Better success rates (*as adjudged by decision-makers*) in their placements across organizational sectors.
9. Fewer failures among, or loss of the 'chosen ones'.

7

The point here is that these expectations are able to be formally evaluated over time. The client, is required to determine the optimal duration where they would expect to observe noticeable 'effects', in the context of their own organization.

Evaluation can be via 360 ratings, custom-designed questions (*targeted to the evaluation targets, not always generic*), financials (*where appropriate*), and judgements made by relevant 'decision-makers'.

Evaluation is likely multi-attribute, a mix of objective quantitative and non-quantitative evidence and subjective judgments, but all focused on providing the decision-maker with the information required to inform a decision about the ROI of the intervention.

The "**R**" in ROI is not necessarily about monetary returns, but the "return" in terms of the formally evaluated expectations of the client. The value-validity judgment in many cases will be very similar to that made in a court by a judge ... the weighing of evidence "for and against", rather than calculating a simple financial return on a financial investment.

And I haven't forgotten:

- ② Over what time-period do you expect it to occur?
- ③ What, for you, will constitute failure?

The second is taken into account in consultation with the client, when the formal evaluation strategy is being developed.

The third is really a counterpart to the first question, because in many cases it is natural to think what success looks like, but sometimes asking what would constitute failure can sharpen the evaluation strategy in terms of proactive evaluation of trajectory patterns which can be a signal for likely future failure, rather than waiting to evaluate at some 'end-point' when any form of proactive intervention is impossible (e.g. turnover statistics).

What if there are no benchmarks?

What if there is no 'history', no information as to previous attempts at solving the problem, success-rates, or employee activity? The client is deploying something for the very first time?

Answer:

The client has to take the punt, while designing and deploying the evidence-base information-accrual strategy that will be used to conduct future evaluation over one or more durations considered feasible, practical, and sensible.

8

Sometimes we have to accept that there just is no 'history' a client can look at for comparative purposes; so they are starting a deployment '*tabula rasa*'. But, the "*I want to know what I'm buying does what it says on the box*" mind-set is the same ... information now has to be collected with that 'down-the-line' evaluation in mind. Clearly, this whole "*should I evaluate?*" question reflects a particular **mind-set** of a professional.

For many clients, to evaluate (or not) is a tricky decision. If the intervention can be shown to be a success, then "*hello annual bonus*"! But what if it's claimed effects have no organizational impact, or even a negative one; how does the client justify their purchase decision? This really goes to the heart of the matter. Psychological interventions are messy, imprecise, and difficult to evaluate, so with the best will in the world, any purchase and deployment will always involve risk. But, with an evaluation mind-set, the expensive ones are always monitored and formally evaluated against expectations. In short, that mind-set is about '**due diligence**'. You don't promise what you may not be able to deliver, but you can show a critical 3rd-party that you are aware of the problem and have good systems in place to detect failing interventions as soon as practicable. But in a corporate world which only recognizes 'success', and punishes failure, even that admission of "*I have to make a judgement-call about deploying this \$75,000 intervention; I can't promise it will work*" may be impossible to make without huge risk to one's own career.

Can those who sell products and services to clients also act as 'partners in evaluation? Again, tricky. The pressure to sell can in many cases outweigh the requirement to develop the kind of evaluation that a client wishes to propose. That is, a client may be asking more of an intervention than the seller knows is likely to eventuate (*partly because of 'boosted' claims made during the sale process*), but the pressure to close a sale regardless may bias the supposed impartial advice given to a client in order to 'steer' them away from the kind of evaluation that is actually required.

Is any intervention worth this kind of evaluation?

If the cost of deploying a one-shot intervention (coaching, psychometric tests, whatever) is less than about \$15k, then no, it's not worth formally evaluating like this.

The cost of expert consultation, time, logistics, developing and evaluating expectations will likely make the ROI negative.

9

We have to be realistic here. There are many kinds of intervention that cost less than \$10k to deploy, such as specialised coaching, one-off team consultations, employee counselling, seminars, psychometric assessments, and interview support etc. To attempt to set up a formal evaluation model for these would likely cost as much in specialised consultancy and in-house time and logistics (*the involvement of time from senior executives*). In such cases the client is best advised to simply deploy whatever it is they had in mind, and use their own judgement as to the worth or otherwise of that intervention.

The evaluation model we are referring to is for big-cost deployments, like the Gallup Q12, or the £1million (at least) deployment of the MBTI across the entire UK NHS. Here, evaluation costs of even \$100k look sensible, if that £1million or so cost is an ongoing annual expenditure. The reason I created a brief evaluation of the ROI of the Gallup Q12 deployment in Carter Holt Harvey (*back in 2002, before CHH was bought out by private equity*) was that the rollout cost of about \$600,000 across the corporate seemed (to me) more of a hopeful but well-intentioned punt than a rational decision accompanied by a formal financial-focused outcome-expectations model. The 'evidence' presented by the sellers of this assessment was based around the mantra "higher engagement = higher profitability". HR simply took the claims made by the sellers of the Q12 as 'gospel'.

However, a simple quantitative demonstration using the evidence quoted by the sellers of the Q12 was sufficient to answer a simple question for the decision-makers: "*what is the probability of CHH making more profit by attempting to increase their employee engagement vs losing money by doing so?*" The answer:

Probability of an increase in profitability = 0.53

Probability of a decrease in profitability = 0.47

This information was presented to the decision-makers in HR, who then made the judgment call as to whether those odds (1 to 1.13) justified an up-front cost of \$600,000 expenditure. HR thought they did; and proceeded accordingly. A few years later, CHH losses were so large that it was sold off, privatised, and broken up into its profitable constituents. See: <http://www.pbarrett.net/stratpapers/gallup.pdf>

And look at the US corporate Best Buy; a poster-child for the Q12. High engagement scores pre-2010, share-price collapse and over 50 store closures post-2010. For the story and analysis of these data, see: http://www.pbarrett.net/tbv/BV6_Talsim_gone-in-60_minutes.pdf

Isn't this just another gimmick to sell consultancy?

So, it's a "gimmick" to provide an evaluation solution to a client who wants to evaluate the \$150,000 per year they spend on management consultants who provide coaching and psychological input to their 'future leaders'?

You've told us the questions, we can do this ourselves.

Who is the 'we' here?

Existing consultants or HR management who have never attempted anything like this before?

10




What about those banks who end up deploying emotional intelligence and well-being assessments and seminars across their workforce at \$60,000 p.a.? What if they want a solution that is able to evaluate the impact of these seminars on their workforce, beyond *"they are nicer people, they have more insight, they understand people better, they have more well-being, etc"*. What about the Armed Forces, Police, Immigration and Customs services who deploy vast numbers of sometimes expensive psychometric assessments? Is it a gimmick to evaluate whether what has been claimed as *'will work for you'* actually does work as expected?

When some see this presentation, or who have heard about those three questions, their reaction is to say *"thanks for that, we can take it from here"*. Unfortunately, those who say this are invariably those who have never attempted anything like this before in their practice because it has never occurred to them that their own practice/claims made to others requires evaluation. As we have said on a previous slide, this approach to evidence-base construction requires a particular mind-set, a commitment if you like to openness and due diligence on the part of a client to whom claims of 'expected effects' are being made.

When put like this, the flippant remark *"we can do this ourselves"* is recognized as potentially carrying a methodology payload for which few consultants are prepared or skilled. Consider the Gallup work-up; that was just a small "look-see" checking the credibility of claims. What would it have taken to help build the evaluation model with the client for post-deployment? Might we even have factored in market activity, share-price, and sector profitability on top of other 'softer' indicators of the claimed effects of 'engagement'?

If other consultants had been routinely building such evaluation models for clients, then none of this presentation would be at all useful because the clear evidence would be 'out there' that such supposedly 'novel' practices were merely 'standard operating procedures'. Perhaps many consultants would argue that they have thought about doing this, and are capable of generating such evaluation models, but clients showed no interest in evaluating their purchase decisions?

So who is going to do this model-based evaluation?

-  sellers of services/product to clients.
-  the purchasers of the products.
-  a special breed of consultant (*or company*) who sits between sellers and buyers, independent of both. The model might loosely be compared to that of a Queens Counsel/barrister ..someone an organization consults to provide expert but independent advice, assistance, and an empirical evidence-acquisition solution.

11

This "*intervention effect modelling*" potentially opens up a very different career path for a special kind of I/O psychologist or management consultant. It's not impossible that the seller of a product or service, or even the client themselves, could evaluate their own claims or expectations. But, right now, this investigative function looks awkward for those whose first priority is to sell, or assist in selling products or services to clients. And for clients themselves, it is unlikely that many have the time and in-house staff with the necessary expertise to design and build evaluation models without any external-expertise assistance.

It is why we have likened this kind of consultancy as akin to the advice/expertise found within particular kinds of roles within the legal profession.

But many might regard university academics as the source of this kind of work, independent as they are of commercial interests and the need to promote their own products or services. In principle yes, there is good reason to expect this from them. But, they need to have the necessary skill-set to work with a client, building an evaluation model based entirely on the client's expectations (*which may be a mix of quantitative, non-quantitative, and pure subjective judgements*).

And some academics try to dictate to clients their own preferred methods of analysis/evaluation (*which they or their particular field happen to use*), rather than build, create, or innovate what it takes to directly evaluate a client's expectations. Sometimes a standard approach can be all that's required, but for many 'soft-skills' interventions, more innovative approaches are likely required to acquire the necessary information which directly addresses the outcome expectations set up by the client, which can subsequently be used by decision-makers charged with evaluating the validity and utility of the intervention set against its associated expenditure.

In the end, perhaps this all comes down to those two words: "*due diligence*", and how I/O psychologists, practitioners, through to purchasers of psychological interventions view the meaning and relevance of those two words to their professional practice?

Download this presentation as a notes-format pdf from:
http://www.pbarrett.net/presentations/NZ_2014.pdf



Paul Barrett (NZ)

+64-(0)21-415625

paul@cognadev.com

Roché Herbst (Canada)

+1-403-510-9984

info@wehcs.com

