360° Feedback Questionnaire Design Is your application making a business difference?

Fast, Cheap, and Good… pick two. If it's fast and cheap it won't be good. If it's cheap and good, it won't be fast. If it's fast and good, it won't be cheap. Fast, cheap and good … pick two words to live by.

Jim Jarmusch

Cost or Quality?

A few years ago we were asked to submit a proposal for the design and management of a 360° feedback process for a major player in the hospitality industry. This firm had been running a 360° feedback application for its management and professional population, a group of around 400 individuals covering three different levels. The system was running with a technology vendor in which every aspect of the process was automated: self-registration, nomination of feedbackers, reminder emails and the download of reports to the participants.

To the company's horror, the evaluation of the feedback exercise was damning. The survey results highlighted no gains, with no evidence of more informed personal development action planning, meaningful follow on conversations with line managers, or facilitating team debriefs to agree priorities for the work group.[1] The survey comments also identified a big "so what?" factor. More importantly, there was no business metric to indicate any improvements in employee engagement, service responsiveness or productivity.

We were short-listed to make a presentation to senior managers for a refresh to outline how we would work with them to reshape the approach: the positioning, design and implementation of a new application. Reading the body language in the room, most of the executives understood our message. Frustrated by the lack of impact of the current system they were looking for an alternative approach.

Oddly enough it was the HR Director who was unsympathetic to our recommendations for redesign. In his words "We are now paying only £20 for each participant set up and report. How can you justify significantly higher fees?"

We pointed out that £20 per report might be well received by the Finance Director. But had he calculated the overall cost of this exercise? Take 400 managers who each ask 10 colleagues to complete the feedback questionnaire which takes 15 minutes. The overall organisational cost: 133 person days.[2] And for what outcome?

We didn't win the work. The firm renewed its contract with its existing provider.

If 360° feedback design, implementation and follow through is about budgets to appeal to the Finance Director, there are any number of technology vendors who provide low cost solutions. Alternatively, if the proposition is to reinforce a shift in culture and management priorities, it is difficult to see how a £20 per pop process can make any organisational difference. Worse, it will be a time consuming exercise with counter-productive consequences.

Five years later after the presentation, the company dismissed its CEO after profit warnings. No doubt several factors were involved in the company's decline. But a mind-set of "keep it cheap" in 360° feedback suggests a firm that had the hand brake on while attempting to accelerate. Much smoke from the burning rubber but with no traction to drive forward.

Design Principles

"I strive for two things in design: simplicity and clarity. Great design is born of those two things." Lindo Leader, FedEx

On line technology has made huge gains in the design, implementation and evaluation of assessment, feedback and survey systems, not least in the speed of set up, launch, project management and report back. And on line technology must incorporate the basics of data security and compliance with the measures of GDPR.

This technology now promises an array of functions in dash board reporting, summary metrics and trend analysis. This is assessment technology as providing analytics to shape our understanding of employees and the organisation to guide future decision making.

But is it? Have these efficiency gains in technology been translated into effectiveness, effectiveness as gauged by behavioural change with organisational impact? Has the intervention resulted in higher levels of engagement and productivity, improved collaboration for innovation, and gains in customer service responsiveness? Or is the only outcome, a set of metrics and visual displays?

We argue that many 360° feedback systems are allowing the tail of technology to wag the dog of a successful application. In the search for better outputs, many have forgotten that these outputs are only as meaningful as the quality of the inputs. The dashboard and dials might look impressive. Division A is higher than Division B by 2 percentage points. Or Country Y is lagging behind Country Z.[3] But if the instrumentation is faulty, we are looking at numbers and graphics without a genuine understanding of performance, for individuals, their teams and for the organisation collectively.

This article summarises the key issues in the design of a 360° feedback application to ensure that outputs are based on informed inputs. In previous articles[4], we argue that like most talent management practices, the "sweet spot" of success and impact hinges on: Content, Design, Technology and Process.

This is the principle of "no one thing". It is the combination of multiple factors - with no "weak link" - that drives exceptional outcomes.[5] In the instance of a 360° feedback application this is the integration of:

- **content**: to ask the right mix of questions in the right way of the right colleague group to combine the right level of support and challenge.
- **design**: an engaging user experience and a report back that is less about the number crunching of standard deviations (or even worse, percentiles) and more about simple visuals for insight.
- **technology**: to provide a secure platform for on line completion, efficient system administration, automated reporting and compliance with data privacy standards.
- **process**: the positioning of the process to create the right "mood music", clarify the rules of engagement and the follow through of debrief and coaching to build momentum and action the key priorities from the feedback.[6]

Design Principles

“You have to start with the customer experience and work your way back to technology.” Steve Jobs

In our experience, few vendors understand the dynamics of the four factors and how to combine them within a coherent application. Some vendors specialise in technology but with little insight into the construction of a robust and credible tool.[7]

Others adopt the strategy of psychometric “precision”[8], an approach that results in windy and repetitive assessments that are time-consuming for the end user.

A few vendors specialise in implementation and the follow through of coaching and development programmes. For these providers, any feedback tool will do. The argument here is that the value lies in the “conversation” with individuals. Of course, facilitation is a key part of the process. But if the tool is flawed, the noise of inaccuracy will drown out the signal of feedback insight.

And in this crowded market place, some vendors seem to lack any appreciation of the realities of human nature, social dynamics and organisational life. These are the systems - insensitive to the challenges of positioning, implementation and evaluation - that become a 360° feedback exercise as “spray and pray” with the hope that the intervention will somehow generate improvements - individually and organisationally.

Design Principles

"Design is not just what it looks like and feels like. Design is how it works." Steve Jobs

The conventional approach has been to construct a questionnaire for completion by participants themselves and their colleagues. This maps out a set of behavioural statements against some "success framework" - often a mix of values and competencies - with a rating scale, with several prompts for open ended comments.

There are other design options. One of our clients asked us to design an application base on a series of images in which colleagues selected the image that best reflected their perception of the individual. This is 360° feedback as a kind of projective assessment in which the choice of images provides insight into the individual, their impact and operating style. As an alternative approach it has the virtue of originality and user engagement. But as with other projective tests, consistent and reliable interpretation of the results is problematic.

If the majority of colleagues have selected a shark rather than say a dolphin for the individual, we suppose they are getting some kind of message about their operating style. But what?

Is it to be less aggressive? Or are they being admired for the ability to keep moving with purpose? Or are colleagues noting how the individual addresses those low-performing employees who are damaging the organisational eco-system?

This article focuses on the construction of a 360° feedback application based on the questionnaire format to address three issues:

- what do "good" questionnaire statements look like?
- which rating scale and response anchors should be used?
- which questions should be targeted at different colleague groups?

What do “good” questionnaire statements look like?

“Try to leave out the part that readers tend to skip.” Elmore Leonard’s Rules of Writing

At one level, the basics are straightforward. Questionnaire statements should be concise, clear without any ambiguity of interpretation and make logical sense. This is rarely the case as one researcher identified in many commonly used questionnaires.[9]

This reflects our own experience in conducting market research and competitor intelligence across a range of 360° feedback products. A surprisingly high number of feedback applications incorporate long-winded statements, faulty grammar and logical inconsistencies.[10]

The criteria of a “good” 360° feedback statement:

Observable

“Strives to improve own performance” requires the respondent to second guess the individual’s intentions or motivation. The individual may or may not “strive”, but to respond to this item requires the art of mind reading.

Similarly, “Is generous in sharing ideas and informs others of new projects and opportunities” asks the respondent to make an evaluation of an attribute - generosity. The individual may or may not be generous; the focus should be on observable impact not the trait of generosity.

Specific and concrete

This is to avoid the generalised statements of abstract vagueness and to write statements that are grounded in specifics.

“Communicates well” seems too broad-brush to encourage useful feedback. “Provides briefings and updates to keep me well informed of any planned developments which will affect me” is a more informed evaluation. For the respondent it provides clarity. For the participant it is more likely to generate the practical insight that can be actioned.[11]

What do “good” questionnaire statements look like?

One thing at a time

Double barrelled statements that combine two different activities confuse.

“Builds relationships effectively and creates a sense of team spirit” asks the respondent to provide feedback on relationship building and team motivation. In this scenario, peers have to decide which element to emphasise. Relationships with the individual might be positive, but do peers have awareness of team spirit within the individual’s work group? Conversely, team members might report that the individual is effective in managing the work group but have less appreciation into the individual’s wider relationships with other colleagues.

Personalisation

“Encourages team members to share openly any problems and difficulties” is not a bad feedback item. But it is probably best asked of the individual’s line manager. For team members it may be better to reframe the question as: “Encourages me to share openly any problems and difficulties”. This is feedback as immediacy. What is the impact of this activity on team members personally?

In generating questionnaire statements, the rule of thumb is to review the item through the eyes of the different feedback groups. As line managers, what does this look like and mean for them? For team members, is this activity helping or hindering their personal performance?[12]

Differentiation

If the vast majority of participants are consistently receiving high ratings, the chances are that the tool is flawed.

Some vendors attempt to get around this problem with a few sleights of hand. Statistically, adjustments are applied in the scoring based on thresholds and weightings. Here mysterious black box algorithms transform respondent inputs into recalibrated scores. This is 360° feedback as might be designed by the Wizard of Oz. At first sight impressive, but when the curtain is pulled to the side, the deception is revealed. When a scoring system can’t be explained to a participant in less than 15 seconds, it is an indefensible system.

Visually, the wheeze applies “cropping” to truncate the scaling in the reports.[13] For example, a 1 - 5 scale is displayed as a 3 - 5 point scale. This has the effect of exaggerating the feedback pattern to make relatively modest differences look much larger than they are. This optical illusion presents feedback that is out of kilter with the intentions of those providing the questionnaire inputs

It is true that well designed feedback questionnaires comprise a spectrum of support to challenge to ensure all participants receive a mix of positive and more “difficult” messages.[14] However, if the majority of feedback reports are all “much of a muchness” that lack differentiation (within and across participants), either the questionnaire is faulty or respondents lack the maturity to provide discerning feedback. In either scenario, there is more noise than signal in the feedback reports.

Which rating scale and response anchors should be used?

"I never understood the frequency, uh-huh" R.E.M.

Historically, frequency scales have often been deployed in 360° feedback construction, along the lines of: 1: Almost never; 2: Not often; 3: Often; 4: Very often; 5: Nearly always

This scale seems to have the virtue of an observable evaluation, without requiring any judgement from colleagues. This is 360° feedback the Roy Walker way: "say what you see" rather than "evaluate what you think".

But the scale suffers from the "Annie Hall" factor. In the movie, starring Woody Allen as Alvy Singer and Diane Keaton as Annie Hall, each visit their own psychiatrist to discuss their relationship.[15]

ALVY'S PSYCHIATRIST: How often do you sleep together?

ALVY: Hardly ever. Maybe three times a week.

ANNIE'S PSYCHIATRIST: Do you have sex often?

ANNIE: Constantly! I'd say three times a week.

This is the definitional problem; frequency anchors can be interpreted in very different ways by different respondents. And, what is the practical difference between "almost never" and not often"?[16]

The additional problem with this scale is that it assumes all behaviours are equally frequent. Two statements:

"Responds with urgency and speed to the issues which others pass for their attention"
"Conducts a "root and branch" review of our organisational infrastructure to identify how and where we need to change."

These two activities seem to be of a very different order.

Responsiveness seems a fairly universal theme within organisational life. Here a check on frequency of activity may be sensible. In contrast, it seems unlikely that root and branch reviews are a regular event. And if the answer is "nearly always" this seems to highlight a liability rather than an asset.

Which rating scale and response anchors should be used?

There is no shortage of other scales to deploy: Expectations, Performance, Effectiveness, Strengths, Improvement Scope and more, with the default position being the somewhat bland Agreement scale.

Behaviourally anchored scales were one attempt at greater precision. Here each response option was accompanied with a detailed description to provide greater clarity for respondents. In principle, a promising approach. In practice, drafting the text for meaningful differentiation across the scale became a time-consuming process in design. And for the respondents, this format became a long-winded exercise in logical reasoning rather than help make rapid evaluations.

Others - in the attempt to mitigate the halo effect - deployed an ipsative response task. The forced choice design in which respondents must select, for example, most and least like from several options has been controversial in self assessment.[17] For those providing feedback it proved to be a highly frustrating and unpopular response task.

This is now a design format only deployed by the bravest - or most reckless - of vendors.

Which rating scale and response anchors should be used?

In a highly thought provoking blog post, Brent Roberts asked the question: "Yes or No 2.0: Are Likert scales always preferable to dichotomous rating scales?"[18]

There has been much debate over rating scales. Is a 1 - 4 scale better than 1 - 5 scale? Or should we extend the scale to 7 points for more accuracy? Or will a 10 point scale provide even more precision?

Is a 2 point - Yes or No - scale a better alternative? It has the virtue of simplicity and speed. No need for respondents to agonise over the difference between "tend to disagree" versus "disagree". And this response task focuses respondent minds on what is and is not working for the participant.

We apply this format in questionnaire design to incorporate the contra-indicators of working life. This follows the observation of Peter Drucker that before we start the excellent we should stop the counter-productive. In this design colleagues, presented with a series of blocker statements, are asked "Is this a potential barrier to the individual's effectiveness?" Yes or No? The number of colleagues endorsing a specific blocker for the participant is then calculated as a percentage of the total feedback group to identify any "red flags".

In our analysis of "contra indicators" we have identified two different themes. The first is the **absence of a positive**; a gap that that is a constraint to effectiveness.

"Focusing only on their part of the organisation and neglecting key interfaces with other work areas." It is understandable why the participant's priorities might centre on their immediate work group. But this gap may have the consequence of silo thinking which holds back collaboration across the business.

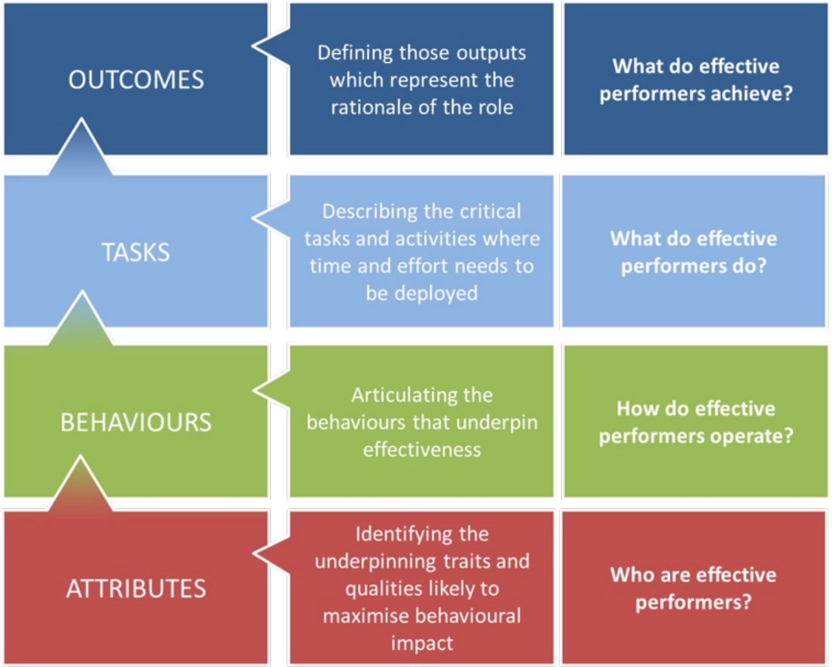
The second type of contra-indicator is the **presence of a negative**. Often this is the risk of over-playing a strength.[19]

"Making unsatisfactory compromises in accommodating competing stakeholder expectations" may be the outcome of an agreeable style that looks to achieve harmony. It also a style that avoids constructive challenge in managing conflict.

The choice of scale obviously depends on the purpose of the feedback application. If 360° feedback is positioned as a process for evaluation to be integrated within an appraisal process, the rating scale and associated response anchors will be different to those systems designed to support development planning as part of say a coaching programme.

Which rating scale and response anchors should be used?

In our model of performance, we identify four levels.



A process built on **outcomes** enables feedback on issues of direct organisational relevance. Clear feedback on outcomes - the "what" of fundamental business purpose - leaves little wriggle room and can be an important motivator for personal change. But outcomes (productivity, service responsiveness, innovation and so on) are a consequence not simply of personal effectiveness but of any number of contextual factors (the legacy of the past, the impact of current situational factors) which may have a favourable or unfavourable effect on outcomes. In addition, respondents, particularly peers and team members, may feel uncomfortable providing feedback which has associations with performance appraisal.

A framework based on **tasks** should provide insight about the effectiveness with which time and energy are deployed across the critical activities that drive positive outcomes. At best, this is a helpful perspective in understanding why outcomes are or aren't being achieved. However, for complex organisations with many different roles and associated task requirements, feedback has to be highly targeted and customised for different roles or fall back on generic tasks that may lack immediate relevance to participants.

Some frameworks focus more on **behaviours** to clarify the "how" of performance. If the aim of the feedback programme is to reinforce cultural values and gain a better insight into operating style and fit, then a blue print of behaviours is the guiding principle for design. The downside of this approach is that the feedback - because there is less direct read across to bottom line outcomes - can lack edge.

Attributes, although often applied in 360° feedback systems, generally should be avoided in design. Attributes require feedbackers to evaluate cognitive and personality traits and motivations, a response task that can result in a halo effect based on "likeability" and lack the specific insight to drive personal change.

There is of course no perfect scale or set of response anchors. Some however seem marginally better (less worse) than others. But 360° feedback, whatever the claim, is not the science of precision. Questionnaire design is therefore more a craft to manage the art of the possible, accepting the messy realities of human psychology, social interactions and organisational politics to balance the trade-offs between credible coverage and user completion times.

Which questions should be targeted at different groups?

Largely for the convenience of the vendors, a typical feedback questionnaire adopts a "one size fits all" approach. There may be different versions targeted at different organisational levels and functions, but for the most part a standard questionnaire is sent to all the different respondent groups, the individual themselves, line managers, peers, team members, other stakeholders, and so on.

Possibly an efficient tactic if the feedback questionnaire is short and focused, say 5 - 10 statements, it breaks down with longer applications. Well framed content begins with the question: **"who knows what?"** Given the dynamics of the working relationships, up, down and across the organisation, who is best placed to provide feedback on which kinds of issues?

In an analysis of our database of feedback responses we looked at the percentage of "cannot say" across the different categories of those providing feedback. We also looked at levels of inter-rater consistency for all the permutations of groups: line manager vs peers, peers vs team members, team members vs stakeholders, etc.

It became clear from this review that it is pointless asking particular questions of specific feedback groups. Either the levels of "cannot say" were high (in which case why ask that question in the first instance?) or inter-rater correlations were extremely low.[20]

Quality feedback outcomes depend on colleagues being "able and willing" to provide informed responses. If peers are unable to provide feedback about the individual's effectiveness at delegation, why ask the question? Or if stakeholders are not in a position to make a meaningful comment about an individual's impact in facilitating team briefings, it seems pointless to include this theme in the stakeholder version of the questionnaire.

If the first principle is knowing "**what not to ask"** of any feedback group, the second principle is recognising "**which additional questions to ask"** only of specific colleague groups.

This is to identify the very different expectations of the feedback categories and ask: what do different colleagues expect from the individual?

Which questions should be targeted at different groups?

Line managers dislike surprises. They want to be informed of any potential problems. Progressive line managers also resist upwards delegation; they are busy enough without being pulled into the detail that should be resolved at the next level down. And line managers look to their direct reports for proactive innovation.

Peers, above all, want responsive support, particularly when they are under pressure. Peers also value a "heads up" about any future activity that will impact on their own work group. And peers want to work with colleagues who share information and expertise openly as part of collaboration.

Team members have a different set of priorities. Team members want consistent direction; no mixed messages or confused objectives. They value fairness and equity in the allocation of work and rewards. Team members also look to their managers for guidance and support to help their own development and progression.

Stakeholders - that permutation of respondents who the individual interacts with on a regular basis outside the formal hierarchy - tend to be one step removed from the day to day action. They are less aware of how the individual operates and more alert to the outcomes that are important to them.

Speed of responsiveness is critical as is the resolution of any potential conflict. They want their issues to be understood and they welcome any innovation that can optimise their own effectiveness.

"This is what happens when we design for everyone — we 'dumb things down' to the point that they become useless or inefficient for most people." Leisa Reichelt , Head of Research and Insights, Atlassian

To recognise the dynamics of these different colleague groups is to see 360° feedback outcomes as the "art of juggling" to manage an array of expectations competing for the individual's attention.

Some individuals are highly effective at managing upwards but allow their team to drift. Others are protective of their own team to the frustration of their peers who experience a lack of responsiveness. And some individuals are so alert to the needs of their stakeholders they over-promise, much to the dismay of their team who become overwhelmed by work-loads.

When 360 feedback was administered via "paper and pencil" the strategy of creating multiple versions for the different colleague groups was problematic On line technology makes this is a relatively straightforward process, with the advantage of shorter completion times (no unnecessary questions). It also removes much of the noise for a clearer feedback signal.

Conclusion

“Design is so simple. That’s why it’s so complicated.” Paul Rand

The objective is to ensure that 360° feedback provides important insight, an insight that identifies priorities for change, priorities which are translated into action plans for development, and lead to change with improvements in performance and a positive organisational impact.

A well designed instrument - with relevant content - in itself is not sufficient for an application to translate personal feedback into organisational impact. Design, process and technology are critical.

But if we don’t ask the right questions in the right way of the right people, our inputs won’t give us the outputs we want. When we prioritise outputs with the gee-whiz factor above the design process of questionnaire content, the cart leads the horse. This is the cart that provides misleading analytics and takes us in a completely wrong direction.

Finally, it is worth noting that as we progress out of the pandemic, and organisations rethink their operating model and working arrangements, there will be a shift in the content of feedback applications. Progressive firms are revisiting legacy applications that no longer reflect changing work patterns to build processes with an emphasis on personalisation around the specific activities that incorporate the dynamics of hybrid employment patterns and virtual team work.

Notes

1. In our experience, those who benefit most from a 360° feedback exercise: update their development plan with well-defined commitments; review the results with their line manager; and share the outcomes with their team to agree future priorities and ways of working.

2. These calculations focus minds to avoid lengthy feedback applications. For one organisation we worked with, a reduction in completion times from 20 minutes to 10 minutes (with no loss in measurement power) made a saving of 80 person days.

3. These kind of benchmarking exercises are often misleading. Given differences - functional and cultural - it is difficult to make like for like comparisons. This highlights a feedback paradox. Participants in, for example, a high challenge culture receive "worse" reports than participants in a low challenge culture.

4. "Evidence-based answers to 15 questions about leveraging 360° feedback", Kenneth M. Nowack and Sandra Mashihi; https://www.apa.org/pubs/journals/features/cpb-64-3-157.pdf

"From Feedback to Insight: Translating 360° feedback into Business Gain"; https://amazure.envisialearning.com/wp-content/uploads/2018/09/FromFeedbackTo-Impact_EN.pdf

5. "Lollapalooza and Exceptional Success"; https://amazure.envisialearning.com/wp-content/uploads/2018/09/The-Lollapalooza-Effect-Charlie-Munger-Talent-Assessment_EN.pdf

6. In our evaluation of 360° feedback programmes and the organisational outcomes, speed of response and discipline in the follow through are critical to impact. On line platforms such as Talent Accelerator help maintain through goal setting, tracking and ongoing feedback from colleagues. https://www.talentaccelerator.co.uk/

7. In a review of a 360 feedback process of one vendor, "Cannot Say" responses were scored as zero in the calculation of average ratings. This is taking the absence of evidence as evidence of absence. Unsurprisingly, managers on the receiving end of this racket took exception to their reports.

8. If 360° feedback had been pioneered by market researchers instead of psychometricians, applications would be much shorter, more focused and more insightful. The attempt, for example, to demonstrate high reliability estimates, in particular internal consistency, ended up with "bloated specific" dimensions which were annoyingly repetitive for the end user.

9. "It's What You Ask and How You Ask It: An Itemmetric Analysis of Personality Questionnaires"; https://www.semanticscholar.org/paper/It%E2%80%99s-What-You-Ask-and-How-You-Ask-It%3A-An-Itemmetric-Angleitner-John/4c913b7829ec47c9552bad06d2a0eee62804fe34

10. One issue in design is the process through which the content is finalised. Consultation with stakeholders is a positive. But not every stakeholder is an expert in questionnaire design. Unmanaged the result is an unworkable application.

11. This also reflects research into effective behavioural change. The more specific the development priority, the easier it is to action for improvement. For example, "Atomic Habits", James Clear; https://fourpillarfreedom.com/atomic-habits-by-james-clear/#:~:text=The%20Four%20Laws%20of%20Behavior,without%20consciously%20thinking%20about%20it and BJ Fogg's "Tiny Habits"; https://tinyhabits.com/

Notes

12. This exercise can be frustrating for those organisations who want complete read across between their competency framework and questionnaire design. Typically, however, these models are written as abstract statements in corporate speak, and lack the nuance to reflect working life and interactions for a feedback application.

13. “Honey, I Shrunk the Scale!”; https://nautil.us/issue/19/illusions/five-ways-to-lie-with-charts

14. Our item analysis of our 360° feedback database allows us to determine the “easier” versus the more difficult statements to optimise the mix of support and challenge.

15. Annie Hall; https://youtu.be/O7nPkpdFAic

16. Biodata measures are designed with factual specificity for greater objectivity and verifiability. For example, “how often does your manager review your performance?” The frequency scale would be displayed as: Once a day; 2 - 3 times a week; Once a week; Once a month; Once a year.

17. “Ipsative Tests: Psychometric Properties”, Paul Englert; https://paulenglert.com/ipsative-tests-psychometric-properties/

18. “Yes or No 2.0: Are Likert scales always preferable to dichotomous rating scales?”; https://pigee.wordpress.com/2018/09/07/yes-or-no-2-0-are-likert-scales-always-preferable-to-dichotomous-rating-scales/

19. “Fear Your Strengths: What You Are Best at Could Be Your Biggest Problem”, Robert Kaplan & Robert Kaiser

20. We shouldn’t expect high levels of consistency given the dynamics of different working relationships. But if the correlation coefficients are next to zero, there is problem.