



Polishing the crystal ball

Super Forecasting to overcome 7 challenges in talent assessment

“It’s easy to be a prophet.
You make 25 predictions
and talk about the one
that comes true.”

Theodore Levitt

Prediction, claim and evidence

“Work units in the top quartile in employee engagement outperformed bottom-quartile units by 10% on customer ratings, 22% in profitability, and 21% in productivity.”¹

“SC Wave is the perfect tool to help predict workplace performance/potential. It predicts overall job proficiency at 0.38, and promotability at 0.59.”²

“A typical job interview has a 1% correlation between what you see in an interview and actual job skills. Using ESP raises your predictability value to 60-70%.”³

“At the Losada line, the percent of disconnection has to be 24.24%, the level of kinetic energy has to reach 50%, the gain should be 16.23%, and the expected performance should be .82.”⁴

“Researchers have developed a tool capable of predicting soldier suicide with pin-point precision.”⁵

“Google’s promotion equation showed 90% accuracy for 30% of promotion cases.”⁶

“The Organizational Maturity Rating presents all stakeholders and prospective investors with a rich and predictive picture of sustainable levels of performance.”⁷

“*I never predict anything, and I never will.*”

Paul Gascoigne

Are these claims valid? Do we even know what they mean? What evidence underpins them? And what are the practical implications for organisations?

There is a contradiction within talent management. On the one hand, the claim is of increasing predictive power. We are told we now have an array of assessment tools and analytical techniques to transform the way in which we recruit and develop employees, build teams and manage succession. A permutation of new psychometrics, social sensing technology and Big Data type metrics gives us access to unprecedented levels of predictive power.

On the other hand, the proposal is that organisations are heading rapidly to hell in a hand cart. The industry warns us that employee disengagement is a “worldwide epidemic”⁸, managers are failing at the extraordinary rate of 60%⁹, and trust in leadership is plummeting¹⁰.

If this latter analysis is correct, we are not seeing the impact of the promised predictive power. Alternatively, the accuracy heralded by the new generation of assessment applications and talent analytics of Big Data number crunching is not delivering the predictive benefits.

The talent management business has got itself in a tangle about prediction.

The 7 challenges for prediction

Well designed and implemented assessment systems have a significant impact on business fortunes. But genuine predictive power to inform plans and decisions in selection, retention, progression, team development and succession is being lost in the noise of flawed research, extravagant claims and distorted reporting in the media.

Prediction is key to talent management, but it is also difficult. We add to the difficulty when we attempt to predict what is unpredictable, or fail to optimise our forecasts of what is

predictable.

A wise understanding of prediction will improve our resourcing and development decisions. It will also manage risk and identify future talent vulnerabilities with the potential to damage our business.

Prediction savvy also makes sense of the competing claims of the talent management vendors who promise solutions to improve our forecasting accuracy, claims which may be more “story telling” from flawed research rather than the application of evidence based practice.

The seven challenges for prediction in talent assessment:

1. Defining **success and the outcomes of prediction** is problematic.
2. The dynamics of **cause and consequence** are confused in the search for predictors.
3. The problem of **small and unrepresentative samples** in much research that claims prediction.
4. Prediction in talent management is complex and our **theories of performance** may not reflect this reality.
5. The **self fulfilling prophecy** is a predictive smoke screen.
6. The **jingle jangle fallacy** creates much noise and the predictive signal is weak.
7. The **evaluation of predictive accuracy** which makes claims of validity difficult to evaluate.

Challenge 1: defining success

For some roles, success is obvious and immediate. But for many roles, success is less than obvious. Even over time it can be difficult to work out what success was achieved and the relative role of skill and luck in achieving the outcome¹¹.

“*Much greater care was being given to the development of predictors than to the criterion measures against which they were being validated.*”

Linda Gottfredson

Success is not always obvious

This is the problem of how to determine the outcome of our prediction. When we turn up at the races and place our bets, we are making a prediction: which horse will win. Our predictions might be poor but at least the outcomes are clear.

For talent management, often it is less obvious what defines success, and therefore what is being predicted. Was Fred Goodwin, the former CEO of Royal Bank of Scotland a success when he was voted Forbes Global Businessperson of the Year in 2003? Or as the CEO behind the UK's biggest corporate bail-out to the tune of £24 billion in 2008, a failure?

Of the many paradoxes of prediction, a general principle is that the more important the prediction the worse our accuracy. It may not be as bad as Malcolm Gladwell suggests: “there are certain jobs where almost nothing you can learn about candidates before they start predicts how they'll do once they're hired”. But it is true that some roles are inherently ambiguous and success can only be evaluated with great difficulty over time. The fact that success - genuine and sustainable success of lasting benefit to the

organisation - can be difficult to measure may be one reason why the talent management business has largely abandoned the evaluation of business outcomes over time.

The target keeps shifting

We throw our predictive dart to hit the board. By the time the dart is in flight the board has changed position¹².

For example, when validity research finds that predictive gains can be made in the selection of Personal Financial Advisers using a simple formula from biodata, the process is applied to the selection of future candidates with improvements in sales performance. But when the financial regulator changes the way in which financial products can be sold, the role changes with different metrics of success, and shifting the types of candidates who will be needed in future.

By the time our research has optimised predictive accuracy, the criteria may have changed. No doubt VUCA - Volatility, Uncertainty, Complexity and Ambiguity - has been over hyped. But it represents a challenge for prediction in talent assessment.

“Prediction is very difficult, especially if it's about the future.”

Nils Bohr

Successful prediction changes success

Research identifies the key predictors for team productivity within an organisation's Research & Development function. The research is robust, the predictive model is cross validated and road tested, and applied to the selection of future R & D teams.

3 years later, the research is revisited and there is disappointment that the original predictive model no longer works, and that a new set of predictive factors have emerged. Part of this may be explained by any number of organisational changes affecting team interactions and outcomes. But it may be an inherent problem within prediction; the better our prediction, the more the outcomes change, which in turn create a different logic of cause and effect which requires a new set of predictors.

As Dave Weisbeck comments: *“When you predict correctly and people take actions on the findings your results actually appear to be worse.”*

Challenge 2: cause and effect get mixed up

Another scenario. Research finds that the top 20% of sales people are high on self-confidence. The recruitment blue print is updated to ensure that applicants are assessed on self-confidence, and that only self-confident candidates are short-listed.

It may be that confidence is a predictor of sales effectiveness. Alternatively, confidence is an outcome of sales performance. Here the consequences of success (higher performance builds confidence) are confounded with our theory of the causes of success (confidence results in higher performance).

Data is collected on, for example, employee engagement and business performance, and a correlation is reported, with the claim that engagement is a leading indicator of organisational success. An equally plausible theory would be that business success lifts the organisational mood, and engagement levels increase. The consequence of success is the cause of higher engagement. More likely there is a complex interaction between cause and consequence in which business success drives engagement which in turn underpins further success.

A Time 1 correlation may “waggle its eyebrows suggestively”¹³ but it doesn’t establish causation. A claim of predictive power requires a longitudinal design to assess outcomes in Time 2.

When the consequences of success shape the attribution of the causes of that success, we are not in the game of prediction, we are in the game of story telling to provide a narrative that makes sense of what happened. We are not improving our prediction of the future.

Challenge 3: over-generalisations from research

The problem of the small sample

A consultancy sends out a research report to their clients in the retail sector. Successful store managers share a similar profile, specifically they score high on learning agility. This is welcome news and the retail sector is quick to apply learning agility to the assessment of future store managers.

It may be that cause and consequence are confused again. More likely however an over-generalisation has been made from a small sample of 32 store managers within one retailer. Later analysis conducted with a larger data set of store managers from a broader range of retailers fails to support the initial finding.

Confidence in predictive accuracy depends on large samples to draw meaningful conclusions. Without decent levels of statistical power, claims will be based on locating patterns within randomness¹⁴.

This is a tough challenge for evidence based practice in talent assessment. Sometimes a small sample is the only sample available. But the assessment industry fools itself with generalised claims of predictive accuracy if the research is based on small numbers.

“With any predictive model, you need to have a means to validate that your predictions are valid.”

Dave Weisbeck

Failure is under-sampled

Research to make valid claims of predictive power needs to sample failure not simply success. Looking only at the outcomes of success is no guide to identifying the causes of the success. Rather than over-generalise from a success profile, we need to pinpoint what is it that differentiates success and failure.

Jerker Denrill¹⁵ for example points out that the traits of the successful entrepreneur - persistency to overcome adversity and the ability to persuade others to their plans - may also be the hallmark of the spectacularly unsuccessful entrepreneur.

The problem is that if failure is rarely sampled, we focus on the qualities of the successful. Improvements in accuracy to forecast entrepreneurial success will result from tracking entrepreneurs over time and comparing relative outcomes to pinpoint what differentiates the successful from the failures.

“The predictive validities are not as high as could be expected given the long history of research and development.”

Frederick D. Smith

The goal posts are moved by unrepresentative sampling

A personality test is completed by a group of current job incumbents - a group of call centre workers - and the data is correlated with metrics of productivity and service quality. There are encouraging trends and a selection system is designed to assess applicants for future positions using the test.

After 12 months the organisation notices however that the pattern of applicant data is very different to the incumbent data. In addition there is no business evidence that the introduction of the test has improved productivity and service quality.

Conclusions about the predictive power of a test taken in one context - current incumbents - may not transfer to a high-stakes situation with applicants in which there is a higher likelihood of faking.

The issues of faking in personality testing in high stakes assessment scenarios are complex¹⁶. However the evidence indicates that around 30% of applicants fake, faking reduces predictive validity, and attempts to measure it through lie scales are problematic¹⁷.

When predictive power of practical importance is claimed we should check that the original research reflects the way in which the assessment will be used in a real life application.

Samples are often too small to have the power to detect a genuine trend, and they can be biased by profiling the successful rather than identifying the factors that differentiate success from failure. And if the research sample is different to the target of future applications, then real life predictions will disappoint.

Challenge 4: the “theory” problem

Talent management requires a theory of performance, and the causes that can be expected to generate more of the effects of organisational value. Which factors make it more likely that an individual will be more successful?

Is sustained success¹⁸ largely underpinned by deep-seated psychological attributes, largely genetically determined? Or does success stem from exposure to key life experiences that shape attitudes and values? Is exceptional proficiency driven by deliberate practice and feedback? Or, as Malcolm Gladwell has suggested, circumstance and timing and luck are the dynamics of success.

Our “theory” of performance will guide how we place our predictive bets¹⁹.

If individual psychology and explanations of behaviour and outcomes are complex, add the social and political dynamics of organisational life to the equation. Then throw into the mix the realities of uncertainty and ambiguity of business change and competition, and the predictive task represents a significant challenge.

What kind of theory will guide not only the choice of predictors we can anticipate to be of practical use, but how we integrate them within decision making? Our predictive theory therefore needs to combine the “what” of assessment with “how” the information will be used within a practical context.

“*In theory there is no difference between theory and practice. In practice there is.*”

Yogi Berra

“Achievement involves a complex interaction of many personal and environmental variables that feed off each other in non-linear, mutually reinforcing, and nuanced ways, and that the most complete understanding of the development of performance can only be arrived through an integration of perspectives.”

Scott Barry Kaufman

At best, our theory of performance meets Thurstone’s three criteria of **simplicity** (easily understandable), **accuracy** (providing explanatory and predictive power) and **generalisability** (applicable across different contexts). There are few theories in talent assessment that meet all three criteria to deliver “powerful elegance”²⁰. Instead we have to make trade-offs to juggle competing priorities within a real world application.

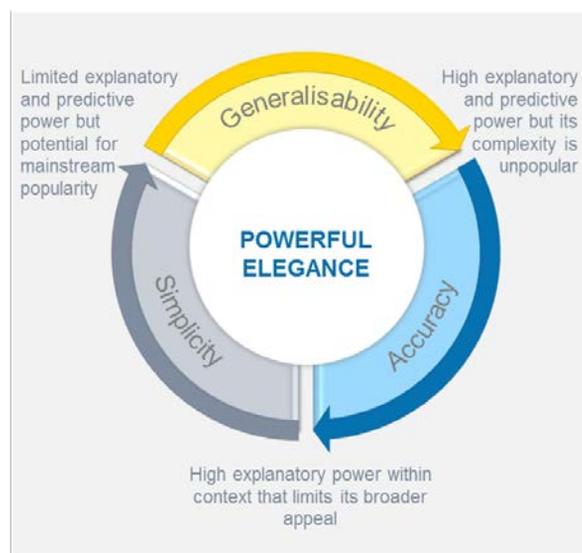
Predictors with simplicity and accuracy but low generalisability give high explanatory power but only within a specific context, a constraint that might limit their broader appeal²¹.

Those with accuracy and generalisability but low simplicity have high explanatory and predictive power but come with a complexity that can be unpopular to practitioners.

Predictors with simplicity and generalisability without accuracy provide limited explanatory and predictive power but are easily understood and applied.

Much of the talent management industry favours simplicity and generalisability. The business logic is obvious; the predictive solution can be packaged easily and marketed with the claim of working “anytime, any place, anywhere”. But in adopting this

strategy, predictive accuracy has been sacrificed for universal appeal. It is also a factor in explaining the “predictive stall” and why prediction in assessment has not significantly improved over the last few decades²².



There is a school of thought that in the world of Big Data, theories of cause and effect are superfluous. We crunch the numbers and look for patterns of association and let these patterns do the heavy lifting of prediction²³. We find, for example, that the browser version used by applicants completing an on line assessment predicts subsequent work success. We don’t need to know why the pattern exists, we only need to know it exists and “works”, and then apply to the algorithm of our selection decision making process.

The statistical reality is that these patterns are often caught in fishing trips in the lake of randomness²⁴ and may be extremely fragile. As Robert Matthews observes: *“Forecasting algorithms give impressive fit to archived data, but can fail badly once they go live.”*²⁵

There is also a practical objection to this kind of predictive strategy. Applying correlational analysis to locate predictors without an understanding of cause and effect may result in processes in talent management with unintended negative consequences;

consequences to keep the Legal Department busy. Talent management without a theory of performance is also unlikely to inform our understanding of the kind of interventions that will improve performance.

The challenge then is applying a theory with sufficient nuance to accommodate the complexities of human nature and organisational life that is understandable and workable within the practical situations it will be used.

“*Just use computer-based regression and correlation analysis to find statistically significant influences, then combine them to get a perfect fit to the data. When a data set is left to speak for itself like this, it typically spouts nonsense.*”

Professor Robert Matthews

Challenge 5: the self-fulfilling prophecy

A new predictor of talent has been announced: the “X Factor”. This factor is advertised to provide additional predictive power to improve existing assessment methods. This “X Factor” is widely reported in the media, and organisations are quick to introduce it into their processes for resourcing and development.

Those individuals highlighted as having the “X Factor” are fast tracked in their career development with access to a range of organisational resources - attendance at the CEO’s Leadership Forum, mentoring from senior executives and participation in a business school programme.

Subsequent research finds that the “X Factor” has no predictive accuracy. Organisations are puzzled. The people who were promoted on the basis of the X Factor assessment are performing well.

This is the self-fulfilling prophecy in which a prediction causes itself to become true. The prediction triggers a new set of consequences (additional organisational support) and associated expectations (people with the X Factor are star performers) that enable the original prediction to become true.

The self fulfilling prophesy incorporates an additional risk in talent management, and may be a key element in explaining the conservatism of much thinking in assessment. If we assume, for example, that CEOs should look and behave in a particular way, or have a particular educational background or career history, our selection decisions will reflect these assumptions²⁶.

But in the absence of genuine evidence, we also may be reinforcing long standing stereotypes.

Rethinking our assumptions about what is and isn’t predictive widens up assessment options. It may also have the positive benefit of advancing greater diversity in the talent pools that can be accessed.

“*The prediction, as it is usual, contributed to its own accomplishment.*”

Edward Gibbon

Challenge 6: the jingle jangle fallacy

When Daniel Goleman announced to the world in 1996 that “compelling research indicates that emotional intelligence is twice as important as IQ plus technical skills for outstanding performance”, this was hailed as a remarkable improvement in predictive power.

The talent management industry saw the opportunity and developed and marketed an array of products based on emotional intelligence at remarkable speed. And organisations were quick to implement the recommendation: select and develop for EQ.

20 years later and there is a significant research base to evaluate the evidence of the predictive power of EQ²⁷. In a recent summary of the meta analysis, after cognitive intelligence has been utilised in assessment, mixed measures of EQ provide an incremental predictive gain of 1%²⁸.

Daniel Goleman himself has now back-pedalled from the initial promise: *“the slow march of research lags far behind the hype of EQ marketers.”*

“*The world of psychological tests jangles more noisily than a gamelan orchestra.*”

Professor Adrian Furnham

The jingle jangle fallacy is a problem of language. When we assume two different predictors are the same because they have the same label, this is **jingle**. And when we think two similar predictors are different because they have different labels, this is **jangle**.

As Adrian Furnham indicates²⁹ this provides much scope for confusion in talent management.

An assessment appears as new. It has an exciting name and it is packaged well. For example, when Grit appeared on the scene, it was jangle that provided the marketing genius with the claim that: “Self-discipline accounts for more than twice as much variance as IQ.”³⁰

Grit sounded different and practitioners assumed it was different. As it turned out Grit was largely a combination of Big 5 personality traits and now looks like it has added little to prediction³¹. It may have even have negative consequences³².

An alternative strategy within talent management is to put jingle at the centre of product development and marketing. This is the strategy of positioning the new assessment application as identical to an established predictor. This is the new assessment that becomes the cuckoo in the talent management nest. It may promise similar predictive gain, but because it is in fact a different measure, it fails in future real life applications.

There is nothing wrong in principle with the jangle of new assessments. Talent management needs to communicate its solutions for assessment and development in ways that resonate with the new challenges organisations face. And a strategy of jingle can provide disruptive innovation by offering cost effective alternatives to the options available from current vendors.

But to avoid confusion in evaluating the claims, both jingle and jangle assessments need to provide evidence of genuine improvements in predictive accuracy over existing measures.

“*At the heart of science is an essential balance between two seemingly contradictory attitudes – an openness to new ideas, no matter how bizarre or counter-intuitive they may be, and the most ruthless sceptical scrutiny of all ideas, old and new. This is how deep truths are winnowed from deep nonsense.*”

Carl Sagan

Challenge 7: evaluating predictive accuracy

Perhaps, the greatest challenge is the sheer fuzziness of what we mean by prediction in talent assessment. For most of us, prediction is simple: we say something will happen and it does.

We expect Don Cossack to win the Cheltenham Gold Cup, and so confident are we in our expectations we even bet on the outcome. And Don Cossack wins.

We also know that nothing in life is certain and we may be wrong. But we know the odds when we place our bets and we can evaluate the outcome. This is straightforward. In talent assessment, it is less obvious what prediction means.

Predictive accuracy is typically reported as a correlation coefficient, the statistic that identifies the relationship between a predictor and an outcome³³. Useful in, for example, meta analysis reviews as an index to compare research findings from different assessment methods, the reporting of validity coefficients has now become a confusing flag waving exercise of claim and counter claim. They now border on the meaningless for talent management practitioners faced with a specific assessment challenge.

Does a move to expressing predictive validity as a probability estimate of success make things clearer?

“*In the era of Big Data we are deluged with false positives.*”

Nate Silver

A hypothetical scenario. A new assessment claims to identify Dark Side³⁴ leadership traits, those factors of narcissism and sociopathy that can be highly damaging to organisations. The web site of the publisher claims the assessment has a predictive accuracy of 70%. Here the publisher references a study in which the assessment identified 70% of a group as Dark Side leaders.

What does this probability estimate mean when applied in a future selection application? To make sense of prediction in talent management we need to recognise that any assessment can get prediction wrong in two ways.

It can predict something will happen and it doesn't. This is the **false positive**. Here for example a diagnostic tool forecasts that a team will succeed; in fact the team implodes after 3 months.

The assessment can fail to predict something that does in fact happen. This is the **false negative**. A Big Data algorithm for example fails to predict the employee who steals from the organisation.

Predictive accuracy requires a knowledge then of the true positive rate and the true negative rate.

“*Wall Street indices predicted nine out of the last five recessions.*”

Paul Samuelson

A data-set is available for 200 leadership candidates that have been tracked for 3 years. This maps out the original assessment results (the data was part of a pilot and not used in selection decision making) and which candidates were predicted to be Dark Side or non-Dark Side leaders. Results from a 360 feedback exercise provide a measure of leadership outcomes to classify the group, Dark Side leaders or not. The full data-set for all 200 individuals is plotted to compare prediction with outcome.

DARK SIDE LEADER	30	70
	False Negative	True Positive
NON-DARK SIDE LEADER	30	70
True Negative	False Positive	
	NON-DARK SIDE LEADER	DARK SIDE LEADER
	PREDICTION	

Of the 140 individuals predicted to be Dark Side leaders, 70 went on to display Dark Side leadership; but 70 didn't. Of the 60 forecast to be non Dark Side leaders, 30 were correctly identified, but 30 went on to be Dark Side leaders.

The predictive accuracy once the false positives are eliminated? 50% , $70/(70+70)$.

If applied in selection the assessment would have ruled out a sub set of non Dark Side candidates - almost a third - from the talent pool; a key issue when competing for specific talent in a competitive market. And this assessment would have resulted in a significant intake (15%) of Dark Side leaders.

Checking false positives isn't enough to establish practical predictive accuracy. To establish predictive power in a real world application, another number is required: the **base rate** of Dark Side Leadership. If every candidate that is assessed is a Dark Side leader, prediction is of course easy. But this is unlikely.

To check genuine predictive power we need to evaluate the prevalence of Dark Side leadership within the target population. If our working assumption is that 5% of candidates are likely to display Dark Side leadership³⁵, how does this affect the 50% rate of "predictive accuracy"?

A larger data-set in another scenario, this time with 1000 candidates is analysed.

DARK SIDE LEADER	15	35
	False Negative	True Positive
NON-DARK SIDE LEADER	665	285
True Negative	False Positive	
	NON-DARK SIDE LEADER	DARK SIDE LEADER
	PREDICTION	

Using natural frequencies, of 1000 candidates, a base rate of 5% means 50 candidates will be Dark Side leaders, and 950 will not. Using the initial publisher claim of 70% predictive accuracy, 35 of the 50 Dark Side Leader candidates will be correctly predicted, but 15 will be missed. Of the 950 non Dark Side Leader candidates, 665 will be assessed correctly but 285 will be incorrectly identified as Dark Side leaders.

The true predictive accuracy of the assessment is 11% , $35/(35+285)$ ³⁶.

“ We don't just want predictions. We want accurate predictions. But it's impossible to judge the accuracy of predictions if we zero in on the hits and ignore the misses. ”

Dan Gardner, Future Babble

The validity coefficient continues to be popular in the talent management industry. Its vagueness allows it to indicate some kind of predictive benefit. But without an understanding of false positives and base rates, it doesn't offer much insight for the practitioner given the assessment challenges they face within a specific context.

The clairvoyant makes predictions, and because these predictions are ambiguous, they can be claimed as “hits”. If predictive claims are not well defined it makes it difficult to evaluate accuracy, and decide which predictive methodology will be more helpful in which context. It also makes direct comparisons of the competing claims from the predictive vendors impossible.

Simple expectancy tables of the kind displayed in this section provide a direct way to understand the predictive gain that can be achieved within a practical application.

“ Statisticians, like artists, have the bad habit of falling in love with their models. ”

George Box

The predictive challenges for talent management

In summary, prediction in talent assessment faces seven key challenges:

- building a better **understanding of what is being predicted**, and how these outcomes impact on business performance.
- **seeing Halo Effect research for what it is**, a poor guide to mapping out the type of cause-effect relationships that give confidence in prediction.
- **distinguishing meaningful claims of prediction** from robust research vs. flawed studies based on small or unrepresentative samples which lack generalisability.
- developing **better theories of performance** to indicate which factors are more or less predictive in which contexts.
- **avoiding the stereotypes of the self fulfilling prophecy** to be more imaginative in where we look for talent and how it can be identified.
- looking for the **signal of genuine prediction within the noise of the jingle jangle fallacy**.
- communicating **greater precision in the claims of predictive power** to make it easier for practitioners to assess the practical benefits in real world applications.

Fortunately, there are reasons to be cheerful.

“*The bottom line is that if you have lots of data and the world isn't changing too much, you can use statistical methods. For questions with more uncertainty, human experts become more important.*”

Lyle Ungar

Recent research drawn from a project funded by a US defence intelligence agency, provides useful insights for talent assessment practice. The story begins with a research programme to track the predictions of experts in their forecasts of political and economic events.

The Super Forecaster Project

In a research programme³⁷ that ran from 1984 to 2004 to evaluate the accuracy of expert predictions, Philip Tetlock was able to answer the question: how accurate are experts in their predictions?

The answer: “dart throwing chimpanzees” would have beaten the experts.

Does this mean that experts aren't any good at prediction, and we should abandon their judgements and pass the prediction task on to statistical modelling? Not quite, some experts performed better than others, and Tetlock wanted to know what made the difference. It wasn't experience or expertise; **it was how they thought.**

Tetlock divided his experts into two groups: Hedgehogs and Foxes. The Hedgehogs are those who know one thing, driven by a single and central concept. Foxes, on the other hand, are sceptical of the one thing and more sensitive to complexity, uncertainty and the nuances of problems. Instead of looking for a single explanation Foxes “stitch together diverse sources of information”.

Whilst the Hedgehogs appeared most confident and credible in their pronouncements³⁸ it was the Foxes that outperformed the Hedgehogs as experts. The Foxes out-predicted the Hedgehogs because they:

- aggregate to **combine multiple sources of information** rather than rely on a single source. Unlike the Hedgehogs who rely on the “One Thing” and aren't interested in finding out more, Foxes search out new information to revise their opinions.
- **think about their thinking.** This is an insight into the typical biases that undermine judgement. It is also Fox type analysis that reflects on conclusions, questioning and checking that their thinking makes sense.
- **display humility** to accept the limits of prediction and to avoid claims of certainty. It is this outlook which accepts complexity and ambiguity and recognises the fallibility of human judgement.

“*Have a prior, collect data, observe the world, update your prior and become a better fox as your work progresses.*”

Nate Silver

Tetlock spotted the beginnings of an approach to improve predictive accuracy. And he found an opportunity to put his ideas to the test.

The Intelligence Advanced Research Projects Activity (IARPA) was a response to the failure by the US intelligence community to assess the situation in Iraq accurately, specifically the prediction of weapons of mass destruction.

The goal of IARPA: to improve the predictive accuracy of the intelligence agencies in their forecasts of key political and economic events. The methodology: a tournament in which different teams would compete in making predictions.

Philip Tetlock rounded up a group of volunteers - the Good Judgment Project³⁹ - to make up one of the teams. And the team performed well, not just “lucky well”, but consistently over time, over 1000s of predictions, to emerge as the predictive winners.

The best predictors in the Good Judgment Project team were identified, the 2% who became known as the Super Forecasters.

“*Forecasting is not some mysterious gift. It is the product of particular ways of thinking, of gathering information, of updating beliefs.*”

Philip Tetlock

Because Tetlock’s team had put the Good Judgment Project members through a battery of tests, it was possible to identify a number of traits associated with the superior predictive accuracy of the Super Forecasters.

Notably they were higher on fluid intelligence and active open mindedness than their less successful peers.

But as Tetlock notes: *“it’s not the crunching power that counts. It’s how you use it”*. What did the Super Forecasters do that helped them make accurate predictions?

The Super Forecasters followed the advice of the Foxes. They:

- kept **predictions in the Goldilocks’ zone of difficulty** where hard work would pay off and avoided the impenetrable issues which couldn’t be resolved. Some predictions are best not made.
- made their predictions **precise**. Rather than rely on vague projections, they expressed their predictions with detailed probability estimates and reported their level of confidence. Apart from allowing accuracy to be evaluated, this encouraged a granularity of thinking about the problems.
- **broke big prediction problems into sub components**, looking at the specifics to distinguish what is known and unknown about the problem.
- **updated their forecasts in light of new evidence** and were prepared to check assumptions and rethink the beliefs that underpinned the prediction. Forecasters who updated their beliefs more often and in smaller increments were more accurate than those who made fewer or larger updates.
- **invested less ego in their forecasts**. The Super Forecasters saw the tournament less about needing to prove their brilliance, and more an opportunity for learning and self improvement.

Improving predictive accuracy in talent assessment

To optimise predictive power in talent assessment, we can learn from the Good Judgment Project and the Super Forecasters and apply five key principles:

- **thinking strategically about the predictive challenge**; here talent assessment isn't simply about probability, but about the consequences of getting our forecasts right or wrong.
- **minimising the need for prediction**; when the predictive odds are not in favour, we should look to other talent management tactics.
- **building a better theory of performance** to have a clearer line of sight between predictors and outcomes.
- looking at **who is making the predictive claim** to combine open mindedness about promising practice with scepticism about the claims of vested interest.
- finding **new predictors**, variations of existing assessment methodology or locating alternatives with the potential to improve assessment validity.

“We can throw numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. Such grandiose visions suggest a failure to understand the limits of brute force computation.”

Professor Robert Matthews

Think strategically about the predictive challenge

Prediction is not simply about probability, and the estimate of accuracy. A strategic approach to prediction addresses **consequences**, and the impact not only of getting our forecasts right, but of getting them wrong.

There are three types of role within an organisation. Those where:

1. Exceptional performance will have a massive and disproportionate impact on business performance

This is the “Lollapalooza effect”⁴⁰; extreme outcomes are produced through the maximisation of multiple success factors interacting in non-linear ways. If we optimise predictive power for these roles, the organisational gains are significant.

The challenge firstly, is to identify these Lollapalooza roles, particularly difficult if the business environment is changing rapidly⁴¹. Secondly, to find the “theory” and predictor mix to provide this kind of accuracy.

If exceptional performance hinges on multiple factors combining in complex ways, this may not be the easiest predictive challenge, but it is one to pursue given the potential business impact.

2. Good is good enough

Some business strategies demand excellence in every area. Most require exceptional performance only in specific areas. For many roles, “good is good enough”. Where the difference in outcome between a good and an excellent performer is relatively modest, gains in predictive accuracy won't make that much difference.

The challenge for these roles may therefore be less about optimising prediction, but finding ways to identify “good is good enough” talent in cost-effective ways. If research, for example, finds that candidates with a specific qualification (that also

commands a salary premium) perform no better than those without, a rethink in recruitment strategy will make significant cost savings.

3. Failure will be a disaster

For some roles, excellent performance won't make much organisational difference. But awful performance will trigger significant organisational loss. Here the predictive focus therefore is less about identifying those candidates who will excel, but screening out those candidates with a high risk of failure.

This is also an easier challenge since predictive accuracy may be better at spotting the failures than identifying the successes⁴².

“*Making better predictions about the future? One way is to limit your tries to areas of competence. If you try to predict the future of everything, you attempt too much.*”

Charlie Munger

Minimise the need for prediction

Charlie Munger of investment powerhouse Berkshire Hathaway makes the point: *“the wise ones bet heavily when the world offers them that opportunity. They bet big when they have the odds. And the rest of the time, they don't. It's just that simple.”*

If certainty isn't on our side, which it rarely is, we should make our predictions carefully, predicting only when either we have to, or when the odds are in our favour. For talent management this is to avoid seeing every talent problem as requiring a predictive solution.

If an organisation is concerned, for example, about the leadership appointments process and the hazard that some leaders will behave in ways that are damaging to the well-being of the business, it can decide to minimise this risk by implementing an assessment of Dark Side Leadership. If the predictive accuracy on offer is only 11%, this is not the best tactic.

An alternative approach would be to look at the strategic, structural and cultural factors⁴³ that either encourage or discourage Dark Side behaviour. This is to shift the theory of performance away from a focus on the individual to see performance systemically and contingent on the context in which employees operate. If predicting which individuals will or won't succeed is problematic, proactive talent management addresses the situational factors and incentives that enable or inhibit success.

We can also improve the accuracy of our talent predictions when we limit the time scales of our forecasts. Asking managers to identify who has the potential to attain Director level in 5 years time - a long-term prediction - may be less useful than the question: which individuals will be ready to advance to the next level in 2017 - a short range forecast.

If line managers are criticised as lacking objectivity in reviewing current effectiveness and impact⁴⁴ it is a big ask to expect accurate

predictions of performance in 2021, particularly if the organisation can be expected to change significantly. Talent reviews and succession plans based on predicting the short term will be more workable and have greater impact than the attempt to look too far ahead.

A variation of the “minimise” strategy is the “fail fast” approach⁴⁵. If we don’t know what will work, and what won’t work in talent assessment, experimentation through pilots may be a better game-plan. Here we move from “predict and control” to “measure and react” to let the results from experiments guide the predictive model.

Randomly controlled trials are the ideal, but are only possible with large numbers. However, for high volume roles even small gains in predictive accuracy will improve levels of performance and retention with significant financial impact.

Rather than embark on a new assessment with the expectation of achieving the predictive power promised by the vendor, a smart strategy would be to run a series of pilots based on assessment variations. This approach builds in a process for robust feedback and evaluates which outcomes indicate predictive accuracy across the different assessment permutations.

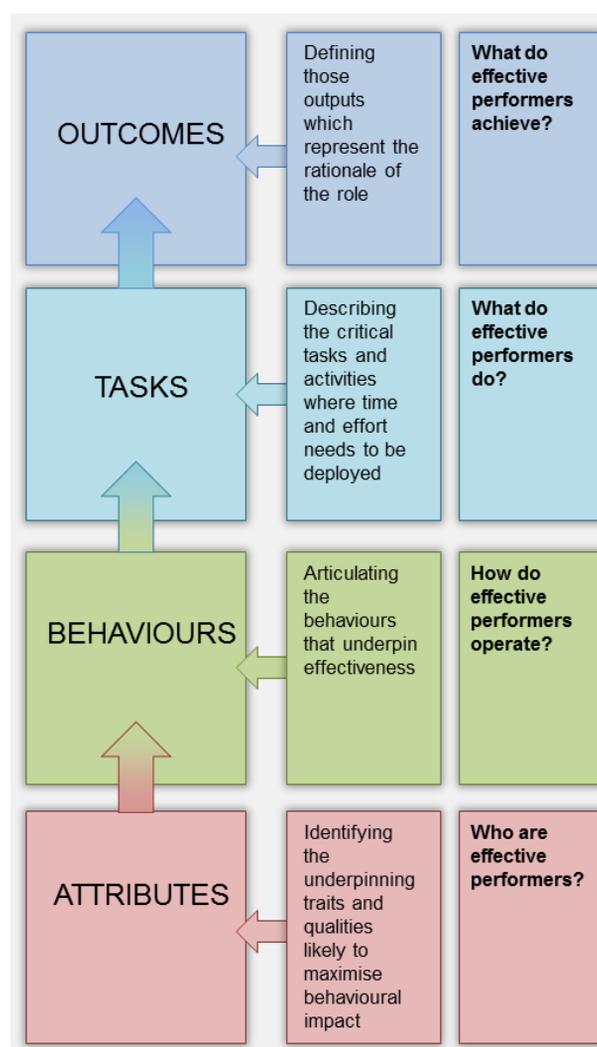
In this scenario - like the IARPA tournament - we are setting up a competition, to determine which predictor mix wins. Historically this has been logistically difficult. A combination of on line assessment, tracking technology and faster statistical analysis now makes the “fail fast” strategy realistic.

“*Nobody cares whether you have a grand theory of success. But are you able to adapt quickly or not?*”

Review of Nate Silver’s, *The Signal and the Noise*

Build a better theory

The widely used “theory” of performance within talent management has been “competency”, typically a framework of between 5 and 15 dimensions to define the success factors important to an organisation. Helpful in many ways as a performance map, competency models can also easily become a “kitchen sink” that bundles together different constructs that make it difficult to apply consistently across different applications (e.g. recruitment and selection, training and development, career progression and succession) or with different groups.



An alternative “theory” is to make explicit the levels of performance that link Attributes to Behaviours to Tasks and then Outcomes. Mapping out this causal chain provides

insight in explaining:

- **how** outcomes are to be achieved
- **why** specific attributes matter in prediction

to provide the potential for more targeted prediction in different talent management applications with different candidate groups.

For prediction, each level has its own distinctive advantages and disadvantages. Understanding these can help improve our “theory” of performance for greater forecasting accuracy.

Outcomes, when defined with precision and based on objective measures, should afford predictive power. If a candidate has a clear track record of achieving the specific outcomes of relevance, predictions of future success can be made with reasonable confidence. Not certainty of course - the outcomes may have been attained through any combination of luck and others’ efforts in a favourable environment. As Warren Buffett remarked: *“A good managerial record can be far more a function of what business boat you get into than it is of how effectively you row.”*

Assessment options include: score card metrics of business impact (e.g. sales, productivity, innovation); management evaluation of performance against business goals, and 360 feedback to evaluate bottom line impact for stakeholders.

A predictive strategy based only on outcomes can however be problematic in two ways. Organisational change shifts the outcomes that are valued. Past achievement become less relevant to a different future. An outcome based approach to prediction may also restrict resourcing options since it limits the talent pool to candidates with direct relevant experience.

“ *All models are wrong but some are useful.* ”

George Box

The evaluation of effectiveness against **critical tasks** is a useful way to track progress against outcomes, and provides insight into the reasons for success or failure.

Assessment options incorporate: portfolios of achievement against key tasks, work simulations and exercises that mirror the challenges of a role, and situational judgement tests and interviewing⁴⁶.

At best, indicators of task attainment should provide predictive power that generalises to other roles with similar task requirements.

The downside: excessive prescription of task requirements can constrain the range of ways in which success can be achieved. And for organisations with a large number of highly distinctive roles, this approach to role profiling can be demanding.

A clear listing of **behaviours** signals how performance is achieved, and articulates the operating styles an organisation values. Here there is a fine line between communicating clear expectations of the “how” of performance and prescribing narrow listings of behaviour that inhibit individuality and limit diversity.

Assessment options include: behavioural interviewing, simulations and exercises, and 360 feedback processes to evaluate behavioural impact.

Utilising behaviours as the focus of prediction typically provides more generalisability across a range of different roles. This approach can also open up imaginative options about talent pools.

However a strategy based on predicting specific business outcomes from behaviours will provide less accuracy.

A predictive strategy based on **attributes**, like behaviours, enables more generalisability across a wider range of roles, but with less read across to outcomes. It can also be expected to have lower predictive accuracy⁴⁷.

Assessment options include: cognitive aptitude tests, personality profiling and other measures of attitudes and motivation, and projective tests. Neuroscience promises the arrival of a new generation of methodologies that tap directly into the biochemical and physiological patterns indicative of key attributes. Whether candidates will respond enthusiastically to these more direct and intrusive assessments remains to be seen.

An assessment strategy based on attributes does however have the advantage of opening up talent pools in resourcing to identify those who could perform in future.

A theory of performance that breaks competency frameworks down into the four levels of performance is a starting point. It provides a better understanding of cause and effect to inform our analysis of which assessment mix will work best given:

- any planned **organisational change** and the impact on the outcomes that are of business importance. If the future is largely more of the same, then outcomes will do a decent predictive job. But if the future is looking very different, this approach may be fragile, and we have to shift to different indicators.
- **organisational structure and design**, and the degree to which there is significant variations across roles, from the spectrum of no roles are alike to most roles are pretty similar. For roles that cannot be defined with much clarity or longevity, attributes and behaviours will have to underpin the predictive strategy.
- the **operating model** and the extent to which culture is largely focussed on the individual and their personal achievements vs building an environment which reinforces collaboration within and across teams.
- the **target talent pool** and the trade-off between the higher cost of proven “outcome” candidates vs the lower cost of the promising “attribute” candidates.

“*When complex adaptive systems obscure cause and effect, make broad and vague predictions.*”

Michael Mauboussin

Look at who is making the prediction

Apart from practitioners themselves, where should we look for prediction in talent assessment? Which claims are more or less credible and provide wisdom to guide which predictors will be more useful in real-world assessment applications?

“*Is the claim true? The only reasonable answer begins with: “It’s complicated.”*”

Tim Harford

There are at least three voices in the talent management predictive business:

1. The vendors selling predictive solutions. These firms - a spectrum covering assessment specialists, test publishers, Big Data analysts and software-based number crunchers - conduct research that provides valuable insights into assessment practice. At best they track the long-term impact of their applications, locate gains in predictive accuracy, and share the findings for progressive practice.

At worst, wrong-headed research is cherry picked and reported selectively, and modest findings are hyped as breakthrough science. This is predictive claim determined by marketing budget rather than evidence based practice.

Here, the Hedgehog - the confident expert who predicts with full certainty - is the hazard. The rule of thumb is: “the more extravagant the claim, the worse the evidence of the claim”. And the “decimal point principle” applies. Any claim based on research that reports more than two decimal points is likely to be the smoke and mirrors of pseudo-science⁴⁸.

2. Academics conducting research and summarising the evidence base from a range of studies⁴⁹. They offer an independent and objective review of what is known and not known as an informed perspective on established and emerging assessment practice.

The potential downside: a combination of impenetrable terminology, and a lack of consensus within the academic community, makes it difficult for practitioners to work out what in fact is being recommended. If the experts can’t agree to communicate an authoritative voice, talent assessment will continue to be fought as a battle of claim and counter-claim.

3. The independent experts active in professional debate and social media. For the most part, these are the progressive thinkers engaged in talent assessment, and alert to new research to summarise key trends in a format that provides insights for improved practice. At best, their credibility and reputation depends on an informed and objective analysis of the issues for end users.

The disadvantage of this perspective can be the overplay of the significance of new findings or distracting controversy resulting from a contrarian attitude. Here it is useful to focus on those who seem most informed and independent-minded⁵⁰ in translating their findings into practical suggestions for better assessment design, implementation and evaluation.

Philip Tetlock makes the point that if the “consumers of forecasting don’t demand evidence of predictive accuracy” we shouldn’t be too surprised that the claims of the Hedgehogs are the loudest.

While the Foxes are alert to the nuances of complexity within context to provide genuine predictive accuracy, the Hedgehogs confidently cut through any uncertainty with simple but wrong claims.

“*Declarations of high confidence tell you that an individual has constructed a coherent story in their mind, not necessarily that the story is true.*”

Philip Tetlock

Find new predictors

The jingle jangle fallacy makes this a difficult challenge within the noise of vendor claim. We need to evaluate the evidence with scepticism to ensure we are not simply adding an expensive measure that overlaps with existing assessment methodology with no incremental predictive gain. Even worse an established assessment is replaced by an inferior alternative.

But the search for new predictors will continue.

This is partly about reviewing the variations in the “usual suspects” of assessment to identify which type of biodata, interview methodology, psychometric tests, work simulations, etc. will work best given the specific nature of the predictive problem we face and our “theory” of performance.

We also need to be open minded to locate predictive gain from emerging trends in assessment research.

Do we, for example, now apply “**snoopology**” metrics? If CEO narcissism can be forecast from the size of photographs in company reports⁵¹, or from the use of pronouns in emails, should we investigate real life indicators that could provide valuable insights into future performance?

Will **projective tests**⁵² make a comeback in talent assessment? Here we accept, for example, that conventional self report measures are proving problematic in high

stakes selection scenarios, and introduce more subtle techniques to understand candidates and the attitudes and motivations associated with superior performance?

Or do we place our predictive bets on **tracking technology**⁵³ to monitor individual and team behaviour, networks and social interactions? In this scenario, assessment becomes a constant evaluation of employees to identify any problems as well as indicators where interventions would result in improvements.

The choice of talent assessment methodology of course is not simply an issue of relative predictive accuracy across the available options.

Our assessment game-plan will be based on the organisation's values and its stance towards candidate and employee privacy and fairness, as well as the practicalities of resource, capability and budget.

“*How predictable something is depends on: what we are trying to predict, how far into the future and under what circumstances.*”

Philip Tetlock

Make prediction work

When we place our bets on the roulette wheel at the casino we have no control over the outcomes. We can only wait for the result.

This is not the case in talent management. In making selection, promotion and succession decisions we don't simply predict and hope for the best. We can optimise our bets by implementing decisions in a way that increases the likelihood of a positive outcome.

Optimising the outcomes from our initial predictions requires attention to a range of additional processes and systems:

- well aligned **induction processes** that bring successful candidates up to speed quickly.
- **feedback** processes that are timely and insightful to indicate what is and isn't working and where improvements will drive performance.
- targeted **learning** applications to accelerate development.
- a shared understanding of **team** processes for collaboration within the work group as well as with other business areas.
- regular **talent reviews** to identify blockages within the pipeline, targets for retention within critical roles as well as the promising individuals who need additional investment to boost their development.
- a **rewards** system that reinforces the right kind of behaviours and outcomes of organisational importance.

Prediction in talent assessment is not a forecast of what might happen. It is a commitment to the future to keep improving the odds that it will happen⁵⁴.

Hedgehog or Fox strategies

Daniel Kahneman⁵⁵ in a recent interview at the Wharton People Analytics Conference noted that some portion of people's job performance is unpredictable and we'll never be able to pick the perfect candidate for every position. This is true but obvious.

He goes on to argue that *“there is no way that by further assessment you can achieve greater accuracy.”*

This is too extreme a position. But it may be a weary response to the hype from the talent assessment industry. The promise here is: through a combination of best practice psychometric testing, Big Data predictive analytics, candidate and employee tracking, the *“layering of different, more sophisticated data will improve the precision of forecasting.”*

There is significant opportunity to optimise our predictive accuracy and improve selection success, raise levels of engagement and retention, and make better appointments in promotion and succession. But unless human nature, social dynamics and organisational life significantly change to become less complex and uncertain, there will be no game changer. Instead, there will be incremental improvement .

And if we limit our options to search for better prediction from Hedgehog claims of a shiny new assessment methodology, we will be disappointed.

Thinking like a Super Forecaster Fox - alert to the nuances of context, applying probabilistic reasoning and thinking systemically about the challenges - will build the predictive strategies that improve talent assessment processes.

“Predicting rain doesn't count. Building arks does.”

Warren Buffett

About AM Azure Consulting

Established in 1994, AM Azure Consulting works with a broad portfolio of clients - in the UK and internationally - in the design and implementation of processes for recruitment and selection; management assessment, development and career management; online leadership tool kits, 360° feedback; performance management and talent and succession management.

If you are interested in our approach to talent management, our consulting expertise, assessment tools and talent planning software:

Call us: 44 (0) 1608 654007

email: officesupport@amazureconsulting.com

Or visit our website www.amazureconsulting.com for further information, including more articles to download for free.

We:

- summarise complexity to provide solutions that are pragmatic and build and maintain momentum for our clients.
- help trouble-shoot the messy organisational problems to see the key issues, identify options and put in place actionable plans that make progress.
- cut to the chase to focus on the distinctive challenges of our clients. We enjoy the innovation that results from our clients with ideas and we help translate them into practical applications.
- draw on an extensive research base, library of resource and range of tool kits, and up-to-date thinking to help design and implement practical solutions quickly.

References

1. <http://www.gallup.com/businessjournal/163130/employee-engagement-drives-growth.aspx>
2. <http://www.psychometricsdirect.co.uk/psychometrics/comparing-the-opq-with-the-saville-consulting-wave-personality-assessments>
3. From the Emergenetics web site, now withdrawn.
4. <http://www.losadalineconsulting.net/#!/losadaline/c5ro> The Magic Ratio That Wasn't; <http://chronicle.com/blogs/percolator/the-magic-ratio-that-wasnt/33279>
5. Martin Seligman on the Comprehensive Soldier Fitness programme, based on the Penn Resiliency Program. <http://anthonow.com/print/throw-a-survey-at-it-questioning-soldier-resilience-in-the-us-army>
6. <http://uk.businessinsider.com/google-promotion-equation-2014-11>
7. <http://www.hrmaternity.com/a-simple-introduction-to-the-maturity-scale/omr-organizational-maturity-rating/> The company that heads this league table was recently fined \$5 billion for fraud.
8. "Leaders must learn to behave better to boost engagement", Advertorial, The Times, March, 2016
9. Personality and the Fate of Organisations, Robert Hogan
10. <http://www.theguardian.com/sustainable-business/2015/jan/21/public-trust-global-business-government-low-decline>
11. The Success Equation, Michael Mauboussin; <https://startuptown.wordpress.com/2015/01/07/success-equation-mauboussin/> Also: https://www.researchgate.net/publication/284559322_Good_Night_and_Good_Luck_Perspectives_on_Luck_in_Management_Scholarship
12. The Texas Sharp Shooter solves this problem by firing at the side of a barn, then painting a target around the biggest cluster of hits and claiming to be an excellent shot. A Monty Python sketch makes a similar point <https://www.youtube.com/watch?v=5lcYIM-47J0&feature=youtu.be>
13. Correlational analysis and the challenges of establishing causation; <http://www.thenewatlantis.com/publications/correlation-causation-and-confusion>
14. John Ioannidis summarises how research based on small sample sizes lacks the statistical power to support the claims they make, explaining why when subject to the rigour of replication - the initial finding vanishes. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/>
15. Jerker Denrell on selection bias; <https://hbr.org/2005/04/selection-bias-and-the-perils-of-benchmarking>
16. The debate goes back to the 1950s when William Whyte in "The Organization Man" provided guidelines to cheat on tests.
17. "Personality Testing and the F word" Griffith & Robie, Handbook of Personality at Work, 2013
18. "Ungifted", 2013; Scott Barry Kaufman overviews the range of performance predictors and associated evidence base
19. Forced ranking appraisal, for example, was based on a theory that assumed performance followed a normal distribution; an assumption that has been challenged; <http://www.businessinsider.com/new-study-debunks-idea-that-human-performance-fits-on-a-bell-curve-2012-5?IR=T>
20. A recent review indicates the only serious contender of perfect elegance that combines simplicity, accuracy and generalisability is General Mental Ability; bit.ly/1IC8ZXk
21. Shaffer, J. A. and Postlethwaite, B. E. (2012), A Matter of Context: A Meta-Analytic Investigation of the Relative Validity of Contextualized and Noncontextualized Personality Measures. Personnel Psychology, 65; Even modest levels of contextualisation can improve predictive validity; which is why bespoke assessment may be a better strategy than relying on off-the-shelf applications.
22. The Predictive Stall in Selection, BPS Assessment & Development Matters, Vol 5, No 4, Winter 2013. Assessment Centre validity, for example, may be in decline; bit.ly/1kJNcK9

References

23. <http://www.johnmyleswhite.com/notebook/2010/10/01/three-quarter-truths-correlation-is-not-causation/>; also, http://www.academia.edu/15170878/Could_Big_Data_Be_the_End_of_Theory_in_Science_A_Few_Remarks_on_the_Epistemology_of_Data-Driven_Science
24. Michael Raynor reviews the business success genre to highlight how easily we are fooled by randomness; <https://hbr.org/2009/04/are-great-companies-just-lucky>
25. "Chancing It" Robert Matthews, 2016
26. The Look of a Leader; <http://www.economist.com/news/business/21620197-getting-top-much-do-how-you-look-what-you-achieve-look-leader>
27. A Critique of Emotional Intelligence: What Are the Problems and How Can They Be Fixed?. Personnel Psychology, 60: 235–238
28. This is not the argument that emotional intelligence (or preferably character) is not important. But that predictive power needs to be put in context of existing assessments; bit.ly/1IC8ZXk
29. <http://www.thesundaytimes.co.uk/sto/public/article1438833.ece>
30. Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents Angela L. Duckworth and Martin E.P. Seligman Psychological Science December 2005; <http://pss.sagepub.com/content/16/12/939.short>
31. Grit seems to be a variation of low neuroticism and high conscientiousness. In a recent review there is concern that it might not be useful when it matters: in high-stakes selection scenarios; <http://www.npr.org/sections/ed/2015/05/13/405891613/a-key-researcher-says-grit-isnt-ready-for-high-stakes-measures>; <http://www.kcl.ac.uk/ioppn/news/records/2016/February/Grit-adds-little-to-prediction-of-academic-achievement.aspx>
32. <https://news.usc.edu/87122/if-the-going-gets-tough-when-should-the-tough-give-up/>
33. Arthur Jensen outlines the three ways in which predictive validity can be reported. 1. An improvement over chance; this is prediction as forecasting efficiency and how much the predictor beats tossing a coin. 2. An improvement of the odds of success; here we need know the base rate of success to evaluate the level of improvement. 3. An improvement in the outcome that is predicted; this requires a randomly controlled trial to compare the relative outcomes from the predictor group with a control group that hasn't applied the predictor. Also: <http://www.nap.edu/read/1898/chapter/7> And the Brier method for the evaluation of predictive accuracy; Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions, Perspective on Psychological Science, 2015, Vol 10(3)
34. Overview of Dark Side Leadership; The dark side of leadership Management Derailment, Professor Adrian Furnham, University College London
35. There has been much debate about likely levels of psychopathy within the leadership population. Whilst frequencies seem to vary across sectors, the much hyped 10% estimate seems an exaggeration; <http://www.hare.org/comments/comment2.html>
36. Through the eyes of a book maker, the odds for the horse "DarkSideLeadership" have moved from 3:7 (banker bet) to 8:1.
37. Philip Tetlock on expert predictions; <http://longnow.org/seminars/02007/jan/26/why-foxes-are-better-forecasters-than-hedgehogs/>
38. The media prefer the authoritative statements of the Hedgehogs rather than the nuance of the Foxes. Tetlock used Google hits to determine the fame of each of the experts. "The more famous the expert, the worse they did."
39. "Super Forecasting", Philip Tetlock & Dan Gardner
40. <https://25iq.com/2015/12/05/what-does-charlie-munger-mean-when-he-says-that-something-is-a-lollapalooza>

References

41. When, for example, the Formula 1 racing team, Lotus, in the 1970s spotted that aerodynamicists were more important to success than the drivers it gained a significant advantage.
42. <http://www.wired.com/2015/01/growth-science>
43. Roger Martin in "Fixing the Game" suggests alternatives, in particular the need to rethink leadership reward systems; <http://rogerlmartin.com/lets-read/fixing-the-game>
44. "Is performance management broken?"; <http://dupress.com/articles/hc-trends-2014-performance-management>
45. "Adapt", Tim Harford
46. Despite the claim of limited validity, updated meta-analysis indicates the unstructured interview fares reasonably well; bit.ly/1wrVGJI
47. Intelligence may be the exception; bit.ly/1IC8ZXk
48. Too many decimal points? bit.ly/1OM73YD
49. Rob Briner, for example, provides regular reviews of evidence based management in talent management; The Role of Scientific Findings in Evidence Based-HR
50. Paul Thoresen, for example, has drawn up a Twitter list; <https://medium.com/@SurveyGuy2/80-iopsych-pros-to-follow-on-twitter-228ff16dcd2c#.bwbol2brr>
51. <http://asq.sagepub.com/content/52/3/351.abstr> act; More snoopology indicators "Playing Golf, and Other Mistakes CEOs Make" http://www.huffingtonpost.com/adam-grant/playing-golf-and-other-mi_b_5536379.html
52. Projective testing: Historical foundations and uses for human resource management, by Carter, Daniels, Zickar, Human Resource Management Review 23 (2013)
53. For example, <https://hbr.org/2012/04/the-new-science-of-building-great-teams>
54. Google's promotion equation and why it failed; <http://uk.businessinsider.com/google-promotion-equation-2014-11>
55. Interview at the 2016 People Analytics Conference; <http://uk.businessinsider.com/daniel-kahneman-why-you-cant-always-predict-job-performance-2016-4?r=US&IR=T>