

Beware the drift from metric optimisation to metric fixation

By **Ian Jeffreys** and **Liam Hennessy**, *Setanta College, Ireland*

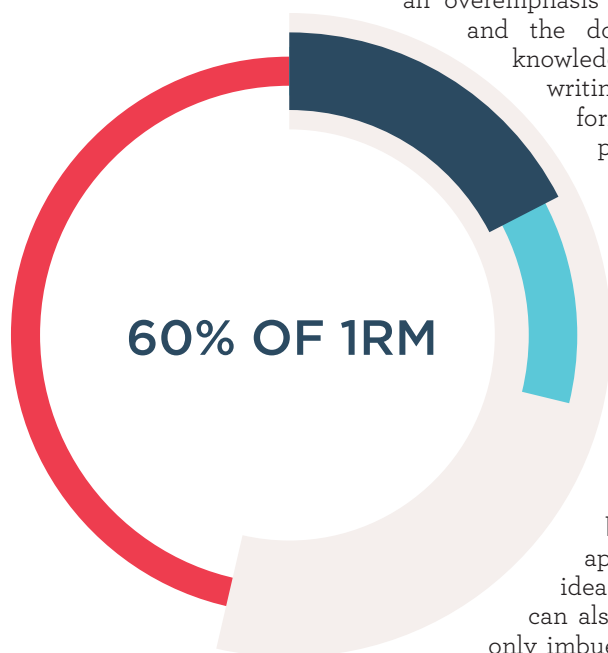
INTRODUCTION

We live in a metrics-dominated world – so much of what we do is measured and strength and conditioning is no exception. In general, metrics are a positive force in strength and conditioning, and used correctly provide a major tool in our armoury – yet they are not without pitfalls. Consequently, although we should be encouraged to utilise metrics, it is also important to remember that their role is simply to help us make decisions – no more and no less. Metrics are not imbued with some magical quality: all metrics have potential strengths, but also potential weaknesses and ultimately metrics have no way of speaking for themselves – instead we speak for them and distil any meaning,¹² and so any value lies in quality interpretation. Effective use of metrics will ultimately involve an appreciation of their value, together with an awareness of their dangers. As Muller⁷ highlights: 'While we are bound to live in an age of measurement, we also live in an age of mismeasurement, over-measurement, misleading measurement, and counter-productive measurement.' The potential benefits and the need to measure are well documented in S&C, yet sometimes we are less aware of the unintended consequences of our use of metrics and so are often blinded to some of these effects. What this article will aim to do is to highlight some of the potential pitfalls of metric fixation and highlight considerations that can ensure that whenever we use measurement and data, the wider considerations are taken into account in order to help us avoid some of the potential pitfalls.

Introduction

In the early days of strength and conditioning, there was a paucity of information to guide decisions. In many instances coaches' decisions were intuitive, based on an iterative process: a method was tried, its effects were observed, subsequent practice was amended and refined. This process then informed what came next, and with each iteration, practice became more effective. However, this process was often held back by a lack of accurate measures via which to evaluate the effects of our interventions. Over time, observations were supplemented

by a range of measures of performance which significantly strengthening this iterative process. However, as these measures became ever more sophisticated and ubiquitous, they may have actually changed the process itself – changing it from a predominantly practical knowledge dominance to a technical knowledge dominance. Rather than being a supplement, something there to help us make decisions, these measures became the main focus, and much of the intuition and judgement that were once dominant were squeezed out of the process. Increasingly, the process of discovery and evaluation took on an external focus, moving



away from practitioner-focused insights to the formal researcher. The iterative process of investigation and evidence generation was usurped by the experiment and the subsequent research paper; the focus on finding individual solutions was replaced by the search for externally validated universal solutions, which in turn led to an overemphasis on technical knowledge and the downplaying of practical knowledge. Today, much of S&C writing alludes to precise formulation, to mechanistic processes, to objective reasoning, to the notion of best practice and the appearance of certainty. Unfortunately, in reality, practice is often far from this ideal and much of what we do relies on practical knowledge and an admission that there will always be uncertainty in any application. Similarly, the idea that data is objective can also be a fallacy, as data is only imbued with meaning when it is analysed, and this analysis will entail a certain degree of subjective bias. To illustrate this, a number of studies across a number of disciplines have shown that even when the same data is analysed by different research groups, quite different results can be elicited, outlining the challenge in achieving a truly objective analysis.¹¹

A universal trend

Strength and conditioning is far from unique in its ever-increasing use of and reliance on metrics. As far back as the nineteenth century, physicist Lord Kelvin is reported to have applied the dictum: 'If you cannot measure it, you cannot improve it.'⁷ Over time this attitude, that measurement lies at the heart of improvement, has pervaded multiple disciplines. In business circles, the American management guru, Tom Peters, embraced the motto: 'What gets measured gets done'; this subsequently became a cornerstone belief of management approaches to multiple disciplines.^{7,9} This type of thinking has pervaded S&C thought, where quotes such as 'if you're not measuring, you're guessing' and 'what you cannot quantify does not exist' have significantly impacted the way we view our discipline. In many instances these approaches helped

develop robust strategies; however, there are also unforeseen consequences of this type of binary thinking. If we believe that if we don't measure, we're guessing, then by design we can be led to believe that by simply measuring then we are entering a world of certainty and taking guesswork out of the equation – yet, whether we like it or not, a certain degree of guesswork is intrinsic in any field that involves human behaviour. Similarly, if we believe that what we cannot quantify does not exist, we naturally believe that what we can quantify is all important. As a result, what we do measure becomes the lightning rod for our thinking processes, to the exclusion of all else. This then plays into the belief that 'what can be measured can be improved' and thereby we step onto the hedonic treadmill of metric fixation.

The creep of metric fixation

Metrics that are strategically planned, contextually appropriate, well-executed and well-interpreted are undoubtedly a force for good, especially where they are seen as being just a part of a wider decision-making process. However, the trends highlighted above generate a momentum where metrics, rather than being part of a wider toolkit, take a dominant and even monopolistic role. Performance can easily become conflated with whatever measurements we may be taking at any given time and our perspectives can become dangerously narrowed. Covertly, our own performance becomes judged, not by our wider impact, but instead by our impact on these measurements; the net result being further narrowing of our focus, even more testing, more documentation, and more reporting – and thereby metric fixation has taken hold. Muller⁷ suggests that one of the main indicators of metric fixation is the belief that it is possible and desirable to replace judgment, acquired by personal experience and talent, with numerical indicators of comparative performance based upon standardised data. It could be argued that we are moving that way. In strength and conditioning another potential indicator of metric fixation is the urge to develop a metric for whatever we are looking at in the misguided belief that without a metric we are unable to analyse and act. Again, it is not difficult to see this trend playing out in many circumstances.

Although there are some aspects of strength and conditioning which lend themselves to quantification, many aspects of human performance do not fit into a neat mechanistic

'If you cannot measure it, you cannot improve it'

framework, and many defy measurement. However, that doesn't mean that they will not affect performance and practice. Similarly, even where we can measure, variables act in combination rather than in isolation and in many situations, we will not know which factors, and especially which connections, are significant.⁴ We cannot expect single metrics to analyse complex situations and lead us to a universal best 'answer'. Indeed, without appreciating some of the challenges of complexity, metric fixation can also move us in a direction that makes us less effective rather than improving practice.⁴

The biggest challenge of metric fixation – the unseen human influence

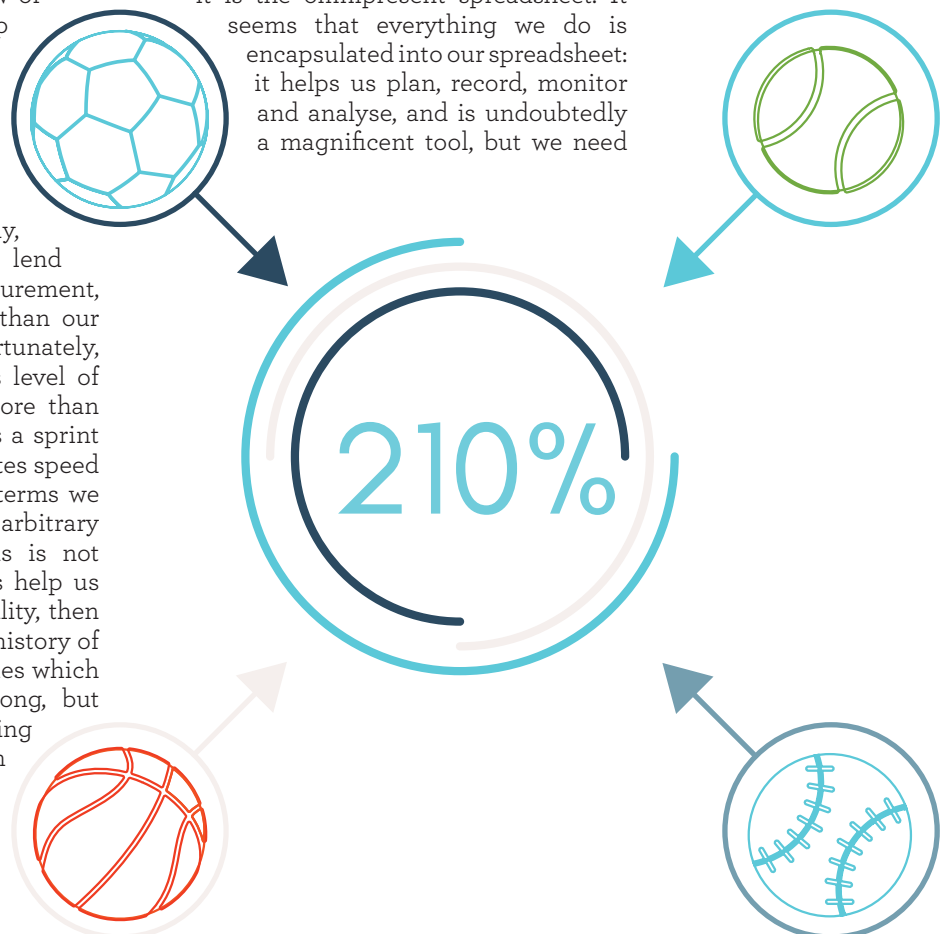
As we have said, human performance is complex, the result of the interaction of multiple factors.⁴ As a result, there will always be an inherent degree of uncertainty and distilling this complexity into metrics is fraught with complications.³ Yet humans are imbued with a distaste for uncertainty, and wherever this exists there will naturally be a pull towards explanation and simplification.³ Metrics play beautifully into this need for certainty and explanation, adding an illusion of certainty to this uncertain world. Once we have a metric in place, it inherently suggests a level of understanding and imbues us with the belief that we are no longer guessing but instead have an objective view of reality. But we should always keep in mind that all of our measures are ultimately human constructs of reality, and not all of them accurately reflect reality. Some of our metrics have long standing scientific validation, our measures of time, of mass etc. Additionally, S&C's truly mechanistic aspects lend themselves to accurate measurement, yet these are far less ubiquitous than our theories would suggest. Unfortunately, not all of our measures have this level of validation, and some are little more than arbitrary values – what constitutes a sprint in GPS data, what bar speed indicates speed strength, and so on. Some of the terms we use are actually little more than arbitrary delineations of performance. This is not necessarily a bad thing: if models help us gain a better understanding of reality, then they can be extremely useful. The history of science is full of models and theories which have ultimately been proven wrong, but have helped develop understanding at certain times.³ Where we run into trouble is where we take these metrics as gospel and fixate on

them rather than understanding both their strengths and – critically – their weaknesses, considering how they relate to the context in which they are generated. In doing this, we give them an allure of certainty and authoritativeness that is often not warranted. This fixation can restrict analysis through the resultant failure to look at a wider range of factors in searching for explanations. This can also exaggerate the underlying belief that reductionist approaches utilising metrics are always the best way to solve all problems.¹⁴ Unfortunately, this approach often sees measures stripped of their context and meaning. Exacerbating this trend is the increasing number of variables we are now able to measure; it is useful to consider that with each abstraction, although potentially enabling a clearer signal to be detected, a considerable amount of noise is introduced into subsequent analysis.¹²

The spreadsheet is not reality

Canadian philosopher Marshall McLuhan has been credited with the quote 'We shape our tools and thereafter our tools shape us'. It is important to consider how our actions are shaped by the tools we use and the environments in which we work. Perhaps the most ubiquitous tool in strength and conditioning today is not related to the mechanics of training at all, but instead it is the omnipresent spreadsheet. It seems that everything we do is encapsulated into our spreadsheet: it helps us plan, record, monitor and analyse, and is undoubtedly a magnificent tool, but we need

'Performance can easily become conflated with whatever measurements we may be taking ... and our perspectives can become dangerously narrowed'



‘Measurement and analyses take time, and metric fixation can see a diversion of resources away from practice and onto monitoring and analysis’

to remember that although spreadsheets are outstanding for managing numbers, they are inefficient at managing intangibles. As a result, we will often over-emphasise that which can be easily measured and almost totally ignore the subtleties of what we do. Quite quickly, and often imperceptibly, the world encompassed by the spreadsheet can become a proxy for reality, again creating the illusion of understanding, an erroneous depth of analysis and an overconfidence in our ability to predict. We start to fall down the rabbit hole of the mechanistic world, leading us to the misguided belief that training is deterministic in nature, and so if we provide the right input, the desired output is inevitable. Unfortunately, apart from a few highly mechanistic activities, this is far from the reality in most activities and there is very little certainty: what we do is far more probabilistic than deterministic.² Yet our spreadsheets can easily lead us into a danger zone where with every level of abstraction we make, our spreadsheets get prettier, but their link back to the actual world is diminished. And so we have a virtual reality by numbers,⁷ yet one that conveniently plays into our need for clarity and certainty. In a spreadsheet, it is almost impossible to duplicate the multiplicities that play out in our training ecosystem. Consequently, no matter how alluring the prospect may be of outlining weeks and even months of training at the press of a button, producing the ‘perfect’ programme, the reality will always be very different and far more complex.



The development of a technocratic hierarchy

Strength and conditioning is ultimately about enhancing physical performance, and effective training has always been and always will be the cornerstone of practice.³ Effectively integrated metrics, as we have seen, can be massively beneficial in this process. Yet more is not always better. Measurement and analyses take time, and metric fixation can see a diversion of resources away from practice and onto monitoring and analysis. This monitoring may throw up flags which for the cautious give reasons not to train; this lack of training may result in a reduced physical capacity, which may throw up further flags to reduce training. The fixation on measurement can easily eat into training time, with time previously spent training now allocated to monitoring. At the same time staff time can be diverted away from

direct coaching into data analysis. To deal with the tsunami of data, new posts may be created simply to administer and analyse data, rather than in employing another coach. Again this may have value, but this value must be considered in the round and on an opportunity cost basis, where what is given up in order to concentrate on metric generation and analysis is fully considered.

The dominance of metrics can become a self-fulfilling prophecy and the huge growth in bureaucracy in many institutions is an indicator of the direction of travel if metric fixation takes a hold. The generation of metrics can give the impression of scientific expertise based on ‘empirical’ and ‘hard’ evidence. This then generates an incentive

to utilise metrics – not only to guide practice, but also to justify positions and to generate an aura of importance. The development and provision of data-driven reports can generate an expectation that reports on metrics are a requirement for effective practice and an ‘interpersonal expectancy effect’ is set up that works along the following lines: middle managers feel the need to generate data-driven reports to their managers, and this then guides the practice of coaches who focus on generating metrics that can be reported, thus often ignoring important, but non-measurable, aspects of performance as these cannot be easily reported; senior managers become used to these reports and then start to expect these type of reports and the process continues. Over time, a managerial culture is set up, which demands more metrics, regardless of its usefulness. Clearly, the use of metrics can help all organisations, but potential problems arise when such measures are not an effective proxy for key considerations and when the generation of data becomes an end in itself, and not a means to an end. Although effective data generation has been crucial to performance in many instances, it is also likely that in others a huge amount of time and effort has been wasted collecting data that has had minimal, if any, impact on decisions.

‘I am not a number – I am a free man’: The human impact of measurement and being measured

The above line from the 1960s television series, *The Prisoner*, spoke to the dehumanising effects of constant surveillance, where identity is reduced to a number. However, when we consider the language of metrics, it is highly technical:

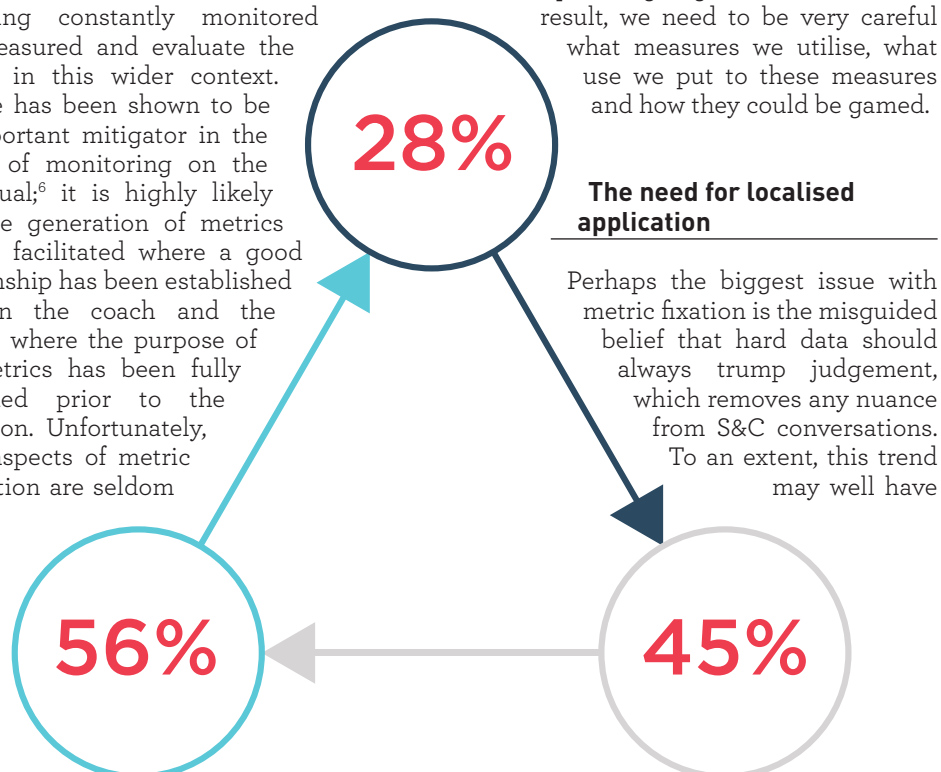
we speak of validity, of reliability of protocols of statistical analysis etc, yet rarely mention any human aspects. We must never forget that our measures are not inert data points, but instead represent the performance of people. Similarly, the collection of data is not exogenous but becomes a part of everyday practice and therefore interacts with all of the other processes present in the system, thus having a direct effect on the system itself. We have already mentioned how the generation of metrics is changing coaches and administrators, but we often fail to consider the athletes being measured and the psychological effects of this constant monitoring. Some athletes may respond positively to measurement, especially where the use of the data is fully explained and where the benefit to their performance is clear and obvious. Indeed, some may even expect measurement, with the use of technology and data being seen an indicator of cutting-edge practice. However, for others, the process of measurement will likely be seen as a dehumanising intrusion, something that may be used against them in performance decisions.

Clearly, the resultant effect of these extreme actions will be quite different, and it is likely that there will be a contextual element to this. Indeed, the results of studies on the effects of monitoring in workplaces have been highly variable and have indicated the potential to both positively and negatively affect a range of psychological variables such as stress, motivation, job satisfaction, trust and commitment.^{5,6} If we are to optimally utilise the potential of metrics, we clearly need to consider the wider effects of being constantly monitored and measured and evaluate the impact in this wider context. Culture has been shown to be an important mitigator in the effects of monitoring on the individual;⁶ it is highly likely that the generation of metrics will be facilitated where a good relationship has been established between the coach and the athlete, where the purpose of the metrics has been fully explained prior to the collection. Unfortunately, these aspects of metric generation are seldom

considered, with considerations normally being restricted to technical analysis. Thus, if metrics are to be optimally utilised, there needs to be greater emphasis on how best to deploy a metric-based approach on a behavioural level and not just on a technical basis.

Another consideration is that the very act of measurement can change behaviour. Whenever anything is measured, and especially if that measure is used as an indicator of performance, there is a risk that the measure will be gamed. British economist Goodhart formulated his law that states: 'any measure used for control is unreliable'.¹ To re-word this in the context of performance means: 'anything that can be measured and used as an indicator of performance will be gamed'. In short, metric fixation invites gaming. Athletes will often have their own incentives to game the system: some may want to game a performance measure to appear as if they have improved on a specific score; others may game a questionnaire to give the results they think we want to see; others may game an initial assessment so that they appear to have made more gains when subsequently tested than is actually the case. Some athletes may even game a monitoring form to avoid having to undertake a certain session. Whatever the reason, the potential for gaming is something that always needs to be considered. Metrics can also direct behaviour, as athletes may chase metrics at the expense of other equally important aspects of performance; it could be argued that: 'once a measure becomes a target, it stops being a good measure'.¹ As a result, we need to be very careful what measures we utilise, what use we put to these measures and how they could be gamed.

'it is highly likely that the generation of metrics will be facilitated where a good relationship has been established between the coach and the athlete'



‘metrics are neither something we must use or must avoid – what matters is their use in context and the value they bring to help us make effective decisions’

been exacerbated by social media – it is difficult to convey nuance in 280 characters. This in turn results in the belief that technical knowledge is the key to solving all problems and the notion that more information will always produce better decisions. Unfortunately, this approach ignores the peculiarity of circumstances and the necessity of practical knowledge to effectively deal with these peculiarities.⁸ In many instances rather than technical knowledge it is practical, local knowledge that is essential and often in these circumstances effective information is that which ‘is as economical and accurate as it needs to be, no more and no less, for addressing the problem at hand’.¹⁰ These decisions themselves require judgement: judgment about whether to measure, what to measure, how to measure, how to evaluate what’s been measured, how best to report and feedback, and this requires more than technical knowledge alone.⁷ This will all be achieved best with local knowledge and metrics that fit the specific scenario, generated by practitioners with local experience. In this way, measurement can never be an alternative to judgment, but must be an adjunct to effective judgement. Indeed, in many instances some of the key factors determining the effectiveness of programmes are the very things on which numerical metrics cannot report: culture, expectations, work ethic and so on. Just because these cannot be measured does not make them any less important.

Revisiting the iterative approach – solving problems

So how should we approach the use of metrics in our practice? Firstly, nuance and balance will be the key; metrics are neither something we must use or must avoid – what matters is their use in context and the value they bring to help us make effective decisions. Ultimately it will be the interaction of theory and practice that brings knowledge and understanding of systems. So interestingly, this brings us full circle to revisit the iterative approach that was fundamental to the development of strength and conditioning practices. Theory and historical practice can guide us towards utilising certain concepts and techniques. Then after each iteration, as we practice and reflect, we amend and refine, eventually updating our beliefs in light of the results we have elicited. As this cycle repeats, over time our practice should become more powerful. Essentially this is a form of Bayesian approach to practice, where we update our beliefs and probabilities based on the results of the previous iteration.¹³ Throughout, this process should be informed both by metrics and our own judgment, with this wider judgment allowing a wider range of considerations to be taken into account.

In deciding whether to utilise metrics and what to measure in this process, consideration should be based on how the metrics can assist in the iterative process.

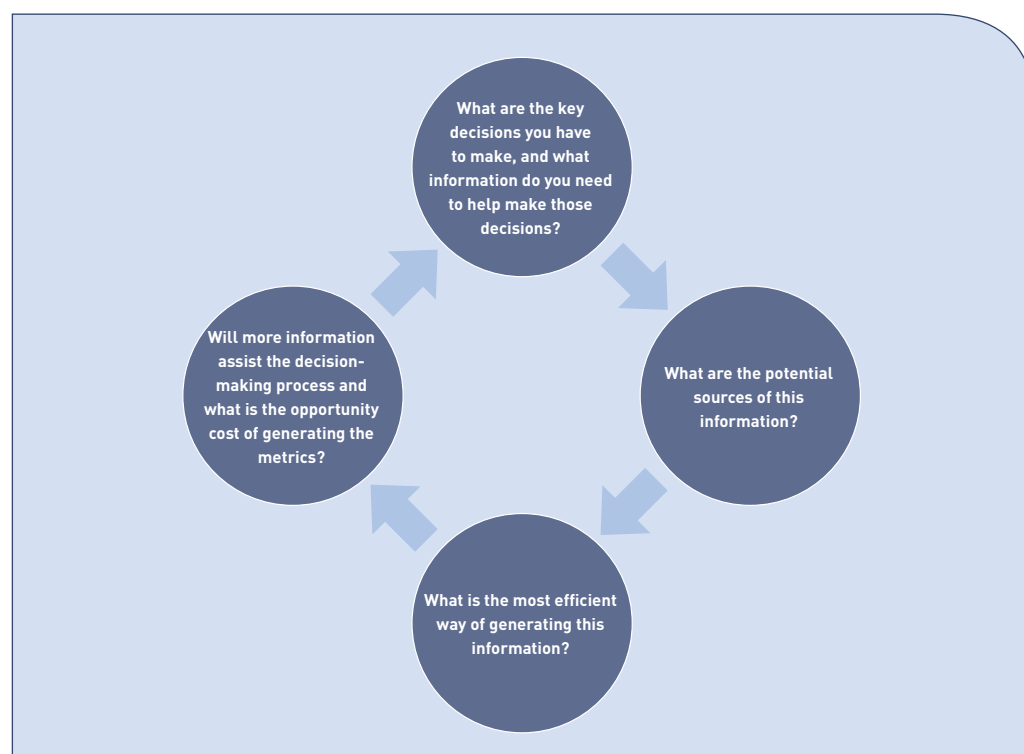


Figure 1. A reasoned approach to the use of metrics

Essentially this can take a reasoned approach utilising the questions outlined in Figure 1.

WHAT ARE THE KEY DECISIONS YOU HAVE TO MAKE, AND WHAT INFORMATION DO YOU NEED TO HELP MAKE THOSE DECISIONS?

These are the cornerstone questions. Ultimately, whether or not to use metrics depends upon what information is needed. Often, the information required depends on the key leverage points that will bring the greatest enhancement of practice, and this clearly relies on context. In some instances, metrics may be of little value, but in others they will be crucial. Similarly, in some situations, very simple metrics will suffice, whereas in others much more sophisticated measures may be needed. Ultimately, there will never be a universal best solution.

WHAT ARE THE POTENTIAL SOURCES OF THIS INFORMATION?

This naturally follows on from the first question: once the information needed has been outlined, a number of sources of that information can be identified. This process will likely throw up a range of quantitative and qualitative approaches and options that range from very simple to highly complicated. Importantly, the key here is not to analyse these but to simply lay out options. All of these will likely have advantages and disadvantages – the key at this stage is not to decide on the preferred option, but instead to simply lay out the potential options. This process prevents options being closed too early and provides for more effective analysis at the subsequent stage.

WHAT IS THE MOST EFFICIENT WAY OF GENERATING THIS INFORMATION?

It is at this stage that the analysis starts. Importantly, this analysis should again relate back to the first question and the function of the potential metrics. It is not about choosing the method that provides the most data or that is theoretically the best; instead, it is about choosing the option that provides accurate information in the most efficient way possible, and which makes the least intrusion into crucial areas of practice. Real life is different to theory and there is usually more than one answer. Considering the option that provides the necessary data in a way that takes the least time, meets the required standards of acceptability, and best facilitates the decision-making processes, will likely lead to the best contextual option. The concept of opportunity costs is crucial here: this relates not just to the monetary cost, but to a consideration of all

of the variables that must be forfeited in taking the chosen option, including time of administration, analysis of what the time and money spent on this option could have been allocated to, and so on.

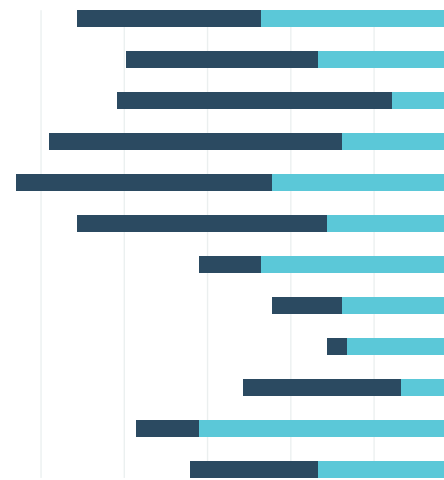
WILL MORE INFORMATION ASSIST THE DECISION-MAKING PROCESS AND WHAT IS THE OPPORTUNITY COST OF GENERATING THE METRICS?

Stage three will normally be the final stage in the initial decision-making process, and guide the first iteration, but – as befits good practice – over time considerations will change as practice evolves and this may necessitate the need for a greater range and/or depth of metrics. With the seemingly never-ending growth in technology, there will always be new opportunities to generate metrics, and ultimately these can also be a natural temptation. However, the key consideration in this decision should be whether more information assists the decision-making process and a further opportunity cost analysis of generating these additional metrics. This analysis should also evaluate the challenge of signal and noise outlined previously, ensuring that if further metrics are explored, they will contribute to more effective decision-making and not simply cloud analysis. It is important to remember that, although we have a great deal of information on correlation, our data on direct causation is far less robust.³

Conclusion

In general, metrics have been a force for good in strength and conditioning, but – as we have seen – they are not without their challenges. Ultimately, the benefits of metrics are not binary but are nuanced, depending on the local situation and the use to which these metrics are put. In deciding upon the best strategy for the use of metrics, it is useful to remember that there are things that can be measured and others which cannot, and it is not the measurement that instils value, but the analysis and its subsequent effect on practice. There will be things that are worth measuring, but that does not mean that everything that can be measured has value. What gets measured may be important, but equally it may have little relationship to what we really want to know, and the critical information may lie in what cannot be measured.

Measurement can enhance practice but sometimes the costs of measuring may be greater than the benefits. Metrics may be able to guide our practice, but at the same



‘Blending the efficacious use of metrics with judgement and insight will bring together the best of both approaches and is probably the best way towards enlightened practice’

time they may distract us from what really matters. When all is said and done, metrics may provide us with a degree of unwarranted certainty that may warp our practice and leave us closed to potential enhancements. Ultimately, we must not stand in awe of metrics and infuse them with a magical value that suggests that without them our practice is flawed. Yet at the same time, we must not fear them and should feel

comfortable in integrating them into our practice whenever they bring potential benefit. Blending the efficacious use of metrics with judgement and insight will bring together the best of both approaches and is probably the best way towards enlightened practice. However, when we get this balance wrong and metric fixation takes over, we can definitely get too much of a good thing.

AUTHORS' BIOGRAPHIES



IAN JEFFREYS, PHD, ASCC, CSCS*D, RSCC*E, FUKSCA, FNCSA

Professor Ian Jeffreys is a coach, author and educator who is currently the academic director of Setanta College in Ireland. A former professional rugby player, he is the author of ten books, a former board member and Vice Chair of the UKSCA, and a former board member and Vice President of the NSCA. He is a fellow of both organisations and the Editor of *Professional Strength and Conditioning*.



LIAM HENNESSY, BA, MSC, PHD

Liam is an exercise physiologist and strength and conditioning coach of international repute. He has coached and worked with some of the world's most prestigious sporting individuals and teams, including professional Golf Major winners, Olympic athletes, the Irish Rugby Football Union, and European professional soccer teams. He competed more than 55 times for Ireland at the pole vault and has served as advisor/consultant to several sporting organisations including World Rugby and Rugby Union. He is President at Setanta College, a leading sport and exercise college specialising in coach education and performance research.

References

1. Arthur, BW. Complexity Economics: Proceedings of the Santa Fe Institute's 2019 Fall Symposium (Dialogues of the Applied Complexity Network). Santa Fe NM. Santa Fe Institute Press. 2020.
2. Jeffreys, I. Evidence based practice in strength and conditioning – reality or fantasy. *Professional Strength and Conditioning*, 39: 7-14. 2015.
3. Jeffreys, I. Effective Coaching in Strength and Conditioning: *Pathways to Superior Performance*. London: Routledge. 2019.
4. Jeffreys, I. How to thrive in a complex world: the role of systems thinking. *Professional Strength and Conditioning*, 60: 33-42. 2021.
5. Jeske, D and Santuzzi, AM. Monitoring what and how: Psychological implications of electronic performance monitoring. *New Technology Work and Employment*, 30(1): 62-78. 2015.
6. Kalishko, T and Riedl R (2021) Electronic Performance Monitoring in the Digital Workplace: Conceptualization, Review of Effects and Moderators, and Future Research Opportunities. *Front Psychol*, 12. 2021.
7. Muller, JZ. *The Tyranny of Metrics*. Princeton NJ: Princeton University Press. 2018.
8. Oakeshott, M. *Rationalism in Politics and other essays*, Indianapolis. Liberty Fund. 1991.
9. Peters, T. *In Search of Excellence*. New York: Harper Row. 1982.
10. Scott, J C. Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed (The Institution for Social and Policy St). Newhaven CT: Yale University Press. 1998.
11. Silberzahn, R, Uhlmann, EL, Martin D et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*. Vol. 1(3) 337-356. 2018.
12. Silver, N. *The Signal and the Noise: The Art and Science of Prediction*. London: Penguin. 2012.
13. Taleb, NN. *Antifragile: Things that Gain from Disorder*. London: Allen Lane. 2012.
14. The Open University. *Systems thinking and practice*. Milton Keynes. The Open University. 2016.