MEASURING AGILE TRANSFORMATION

endava.com/agile

MATT CLOKE
PRINCIPAL CONSULTANT

CONTENTS

AUDIENCE

This whitepaper is aimed at anyone who is involved in making organisations more efficient through the application of Lean and Agile methods for delivering systems. In particular this whitepaper is aimed at coaches, managers and leaders who want to develop teams and maximise their ability to improve.

INTRODUCTION

One of the promises of development and process models that implement Lean and Agile practices is that data collected from the process can be used to tune and improve it. Ultimately, the goal is to demonstrate to stakeholders that a change in approach has increased the speed with which value is delivered to customers. This whitepaper explains the role of metrics in an Agile Transformation, considers the potential pitfalls from applying metrics and then presents a number of metrics that could be applied to teams, sets of teams and to the actual transformation of an organisation.

WHAT ARE METRICS?

Whilst there are many definitions of the word metric, for the purpose of this whitepaper, a metric is defined as a standard of measurement that is used to help evaluate a complex system.

Characteristics of
Good Metrics

1.

METRICS ARE MEASURABLE

A key characteristic of a metric is that it needs to be unambiguously measurable. Whilst there is extensive material available describing how "everything" can be measured, it is important to remember that sometimes the effort involved in making the measurement outweighs the value that it provides.

Most lean and agile development efforts will have some form of development tooling, such as a continuous integration or build environment, as well as tools that organise the work of teams (Jira, Rally, etc.). These tools offer a rich set of data, but to turn them into useable metrics often requires some form of analysis. It is important, therefore, to not just measure the easy things, but to look for the deeper measures that can be drawn from the base data. Furthermore, the metrics need to reflect the progress and health of the Leand and Agile process. For example, an automated build system such as Jenkins can provide information about the frequency of builds, but the value being delivered may be better measured by the duration since the last release to production.

2.

METRICS ARE VISIBLE

When introducing metrics to teams, groups of teams or at an organisational level there is a temptation to guard the metrics and ensure that data is only shared on a need-to-know basis. Along with other Lean and Agile processes it is suggested that openness and collaboration are embraced, and that metrics are not only clearly articulated, but that they are visible to everyone. In this way, teams can be encouraged to understand and embrace metrics.

This visibility can be in the form of constantly updating "information radiators" through to the simple act of posting printouts on walls in communal areas. There should be no opportunity for people engaged in a transformation to claim that they are un-aware of the metrics in use. Confusion over the existence of metrics or what they currently indicate will undermine the value of the measures taken.

As is discussed in greater depth in the Metric Pitfalls section, it is important to consider the Hawthorne Effect when making metrics visible, explain and re-affirm why metrics are most valuable when applied on a regular basis.

3.

METRICS ARE REPEATABLE

Even in cases where analysis is required for a metric (like our CI example earlier) to support the usability of the metric, and its visibility, its collection should be an automated process. Much in the same way that active monitoring helps maintain the health of a system, metrics capture and analysis should be seen as a continual and repeatable process.

Deferring the capture of metrics to manual processes introduces latency in the process as well as increasing the possibility of erroneous data.

Once automation has been put in place, there is the opportunity to scale existing metrics across teams, if required, or conversely to roll out a new metric quickly to replace a metric that is no longer valid. This can be achieved through treating the scripts and processes surrounding data capture as its own platform with all code etc. subject to the same requirements for control within a source code repository and with the capability to automate the build, test and deploy process to support reproducibility.

4.

A METRICS RELEVANCE CHANGES OVER TIME

When looking for metrics that can be applied to a particular subject, it is important to realise that there is a maturity dimension that can be applied to a metric. That is to say a metric can't necessarily be applied universally and some can only be applied, and value derived, when a certain level of maturity or a goal has been met. In some cases, measuring a metric relevant for lower levels of maturity when greater levels of maturity exist can cause false reading to occur.

The maturity dimension means that it is sometimes difficult to compare and contrast metrics across teams, despite the fact that the same metric may be applied. An example of this could be velocity, which can be negatively impacted when an experienced team is asked to solve a difficult problem.

Metric Pitfalls

For those transitioning from a Waterfall development approach to a Lean and Agile development approach, the potential for metric capture is a major benefit. However, a number of pitfalls can emerge from the capture and analysis of metrics in a development process.

MEETING THE METRIC NOT THE GOAL

Observer Effect [2] – "... the fact that simply observing a situation or phenomenon necessarily changes that phenomenon."

When applying metrics in a transparent way and making them visible, there can be a tendency to only focus on meeting the associated goals with the metrics. Historically, this has been seen associated with inappropriate measures such as "lines of code" and "function point analysis", being used to measure the productivity and progress of development teams. These measures have been shown to encourage the wrong behaviour from development teams, with overly verbose code, and complicated function decomposition, just to meet the measure.

It is also the case that when a push is made to meet a specific goal, there is the tendency to let good engineering principles slip and for good practice to be abandoned. Over an extended period of time, this can lead to the "Normalisation of deviance" [1]. This term applies to the gradual erosion of good practice within an organisation and the difficulty of observing it from within.

BEWARE THE SHORT TERM UP LIFT

The Hawthorne effect [3] – "... reactivity in which individuals modify an aspect of their behaviour in response to their awareness of being observed."

It is important that teams are engaged and involved in the creation of, capture and review of their metrics and this can have its own pitfalls. It has been shown that when people know that they are being observed there will nearly always be a short-term change in behaviour.

We've all been in a situation where, when we were the focus of a coach or a trainer, we doubled our efforts, only to ease off when the coach walks away. This effect is known as the Hawthorne Effect and can give coaches a false impression that measuring a particular aspect of a team has a positive effect. Therefore, it is important that both teams and coaches are aware of this potential problem and focus on embedding good behaviours through repetition and feedback.

TEAM METRICS

Lean and Agile methods focus on creating empowered, cross-functional teams, able to deliver change with minimal dependencies on others. Much has been written about the creation of teams and the stages they go through (see "Tuckman's description of forming, storming, norming and performing" [5]) and therefore the metrics described here focus on performing teams rather than measures that can be applied as teams undergo their growth and development.

As with many metrics, there will be a temptation to compare the measures from one team against another, when in fact the true value is understanding the trend over time. It is also suggested that beyond a set of simple core team metrics, the Team themselves develop, measure and refine them.

It is therefore suggested that the following 3 key indicators are observed:

- Health
- Cadence
- Improvement

1. HEALTH

Much like an individual's health, team health indicators and certain 'symptoms' may relate to broader issues and can also indicate when a team requires additional focus from coaches.

MEASURES

Happiness Indicator – It is usual for Scrum masters to "check in" with teams on either a daily basis or as part of a structured retrospective. By simply standardising on a measure, be it 1-5, or sad, ok, happy, it is possible to measure happiness over time and across teams. As with relative estimation, it is more important to consider "happiness over time" as opposed to simply looking at it in isolation. "Is a team becoming happier?", "Is a team reporting un-happiness after a particular change?", are questions that Scrum masters and coaches should be looking at and then work with Teams to address the underlying drivers.

Meeting Count – Particularly when executing Scrum, there is a very formalised set of ceremonies that must be followed. Furthermore, as the principles behind the Agile Manifesto emphasise, face to face conversation is the best form of communication. Therefore, measuring how many meetings are being attended outside of core ceremonies can be a very key indicator of adherence to agile ways of working. There is an engrained culture in many large organisations that work cannot be resolved without meetings, and that it is a frequent anti-pattern for teams to fall back into when transitioning to Scrum. A team with a zero or very low meeting count is probably happier than one bombarded with meetings and external interruptions.

Learning – In any cross-functional team there will be team members with mastery of particular skills. To ensure that teams can operate effectively in the absence of a single team member and to ensure the highest chances of multiple team members being able to complete work items, measure how a team is learning and sharing. There are a number of ways in which learning can be measured, but simply recording on the team's whiteboard when a task has been completed, after a coaching or pair session with another team member, can lead to a useful measure over time. If the measure is recorded over a longer period of time, it can be possible to identify the tutors within the team, those that frequently pair, versus those that don't become as engaged. As with most team measures, these metrics give Scrum masters and coaches the ability to identify coaching opportunities.

2. CADENCE

Agile and Lean methodologies structure work into smaller units, so that it is possible for individuals and teams to observe progress. Seeing change released on a regular basis builds team confidence, releases value for the business and creates a sense of purpose. This regular pattern of work creates a cadence visible to all parties. These regular working patterns and cadence will produce output and over time, this output will begin to be normalised. When this happens, it is possible to match cadence against delivery.

Cadence in sport is important as it describes a correct rhythm at the right point in time (be it the steady stroke rate at the start of a rowing race, or the attack up a steep hill on a bike). Similarly, the cadence of a team is situational and can changed based on where they are in both formation and growth as well as what they are being asked to deliver. It's important to realise that there are a multitude of reasons which can impact a team's cadence. For example, when beginning a new, relatively unknown, piece of work, a team's cadence may drop from a previous level, as difficulties are explored or mastery in a new technology is gained.

When cadence is impacted, and a regular rhythm of delivery isn't occurring, teams begin to feel un-settled and retrospectives usual indicate there are more items to be worked on and resolved.

Some may wonder why a team metric would focus on cadence as opposed to velocity. Velocity measures are focused on the ability to track work done over a given time period. For example, "burn down" is frequently used within a sprint to track how the velocity of the team is delivering the committed story points. The difficulty with measuring velocity is that estimation should belong to the team and should be considered relative with each sprint, not necessarily harmonised or standardised across teams. When transformation exercises attempt to harmonise or standardise estimation across teams, and measure velocity, it is usually in an attempt to revert to tracking teams in a "man-hour" or "man-day" manner and primarily judging when teams have committed to deliver by. By reverting to this form of tracking, there is a danger that too much future extrapolation is applied to future milestones and commitments are made without the involvement of teams.

MEASURES

"Un-done" Work – Scrum teams commit to a product owner to deliver a particular volume of work in a sprint. Many factors can impact their ability to deliver against that commitment, and an un-done work measure reports the proportion of work that was not completed within a sprint. It is suggested that un-done work is measured at the story level (with a record of the estimated story points). This approach will allow teams to consider their experience in future planning. Over time, the amount of un-done work should reach zero. If teams continually fail to complete work in a sprint, then it is likely that the normalisation of deviance has begun, and a coach should look to intervene.

New in Sprint – Scrum suggests that taking new stories into a sprint once it is underway is an anti-pattern. New stories usually suggest that there is either a problem in the planning process (too little work taken into a sprint) or that there are external factors, which are outside of the scope of the team, that are causing teams to lack empowerment and autonomy. In a pragmatic sense, during a transformation to lean and agile practices, there are cases where teams are asked to do more once a sprint has been committed. Therefore, it is suggested that a valuable metric for a team to measure is New in Sprint, and if a long-term pattern is identified coaches should focus on planning to ensure there is either spare capacity within a sprint to compensate for external factors or that further refinement (beyond the suggested 20% of sprint time) occurs prior to planning taking place.
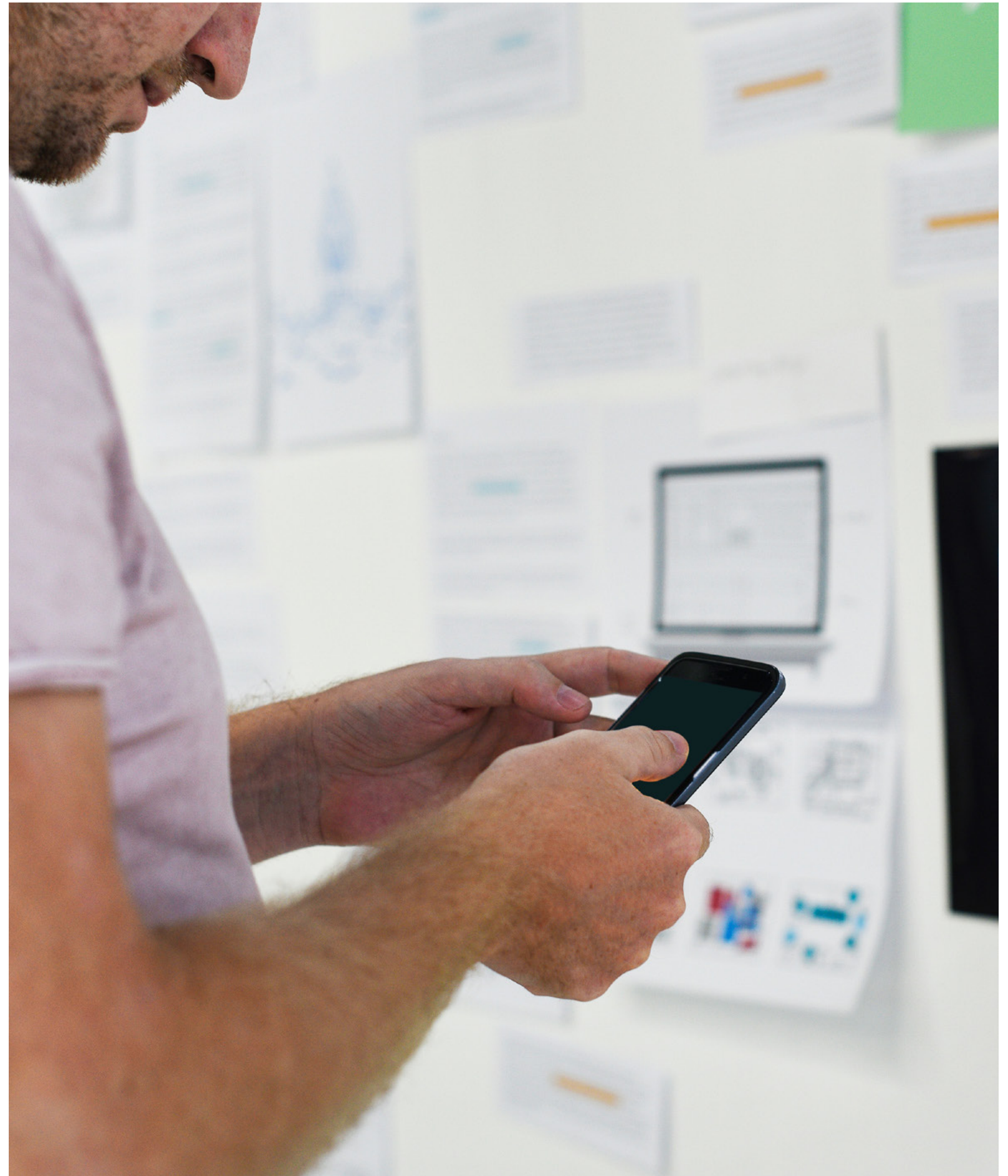
3. IMPROVEMENT

At the core of lean and agile practices is a desire to continually improve through inspection and adaptation. Therefore, it is important that teams use measures that quantify the improvements that are being made. Within Scrum the retrospective ceremony is an important time for a team to reflect on what has happened and what experiments can be conducted in the future to improve the process. In Lean, the concept of the "Andon Cord" [6] is used to illustrate that anyone within a team can alert other "workers" to a quality or process problem and initiate joint work to resolve the problem.

Therefore, it is important that teams measure improvements over time, and unlike some of the measures, this one can be applied to highlight teams who are embracing improvement and teams that require intervention from coaches.

MEASURES

Action items arising from retrospectives – In Scrum the retrospective is a time to identify items from the previous sprint that require addressing in future sprints. These items can include everything from how the team is bonding (do the team eat together, etc.), the number of defects uncovered whilst testing the change, etc. Given that a team cannot focus its entire time on process and quality improvements, the action items arising from a retrospective are the items that the teams are committed to resolve in the upcoming sprint. This count (in conjunction with the closed items, see below) provides a picture of how committed teams are to fixing problems for themselves versus shrugging their shoulders and carrying on.

Closed action items arising from retrospectives – Scrum teams can become habitualised to perform retrospectives, generate useful action items, but never actually address them. Therefore, the closed action items count provides an ability for a coach, and broader transformation team, to see how teams are going about improving and learning from addressing action items.

Metrics Supporting **Scaling**

The scaling challenge in agile occurs when the scope of what needs to be delivered cannot be delivered by a single team. This can occur due to the size of the backlog or factors relating to an underlying architecture or technology. When the scaling challenge is encountered there are a number of agile scaling frameworks that describe how teams can work together and manage dependencies.

Some scaling frameworks (see "SAFe Metrics"[8]) provide a comprehensive set of metrics that can be used when applying the framework. It is suggested that, as with all frameworks, the measures chosen are in line with what is important in the context in which they are being applied, rather than applied wholesale. Furthermore, it is suggested that for frameworks that are less prescriptive about what can be measured, there is value in reviewing those that exist elsewhere.

Given the context of the scaling problem, the metrics can be grouped under the following high-level indicators:

• Dependency Management
• Value Release
• Planning Overhead

DEPENDENCY MANAGEMENT

When initiating a transformation, or when there are specific reasons that legacy complications can't be simplified, there will be a number of dependencies that teams will have on one another or on external teams. To mitigate these issues, and whether using a simple Scrum of Scrums model or a larger Program Increment planning approach, there will be the need to regularly identify and review dependencies. By capturing the number and source of the dependencies it is possible, over time, to build up a picture of constraints within a value stream. This information can give coaches and those involved in a transformation the ability to identify where the effort needs to be applied when simplifying value streams.

MEASURES

Dependent stories per sprint – In any sprint, a team can track the number of stories that have a dependency that can't be controlled or delivered by the team. Even if this is known ahead of time (due to big room planning etc.) it is still a useful metric for a team or coach to track as it can highlight issues with a particular architecture or process. Furthermore, an inability to complete a story due to a dependency can also reveal a problem with the team's agreed definition of "done".

Dependency source – If dependent stories occur frequently, then understanding where the dependency comes from is a useful measure. For example, if Team Penguin is always reliant on Team Walrus, then this could indicate a number of things which could be investigated by coaches, such as skills mixes of teams, physical proximity, poor co-ordination etc.

VALUE RELEASE

When agile teams scale, and dependencies begin to emerge, there is an inherent risk that it takes longer to release value from the process. There's also the case in which, when applying scaling frameworks, the temptation to fall back to habits that may have existed within a waterfall value stream appears. These habits manifest as Program Increments (in SAFe parlance) or Viable Product being available on quarterly or half yearly basis. This is the antithesis of an agile approach and therefore it is important that measures are applied when scaling to ensure value is released on a regular basis.

MEASURES

Time to release - There are multiple ways in which time to release can be measured, but when applied to scaling it, is usually a good idea to measure the release time of the cumulative teams. For example, if Mega Product can't be released until team Penguin, team Olive and team Walrus each individually make their change, then the metric should measure the cumulative time taken. Coaches and teams should then reflect on this time and seek to look at improvements in architecture, process and planning to decrease this measure over time and never forget, the goal of a Scrum team is to release value at the end of every sprint.

Change failure rate - Whilst it is assumed that scaled agile teams are supported by tools that support continuous integration and delivery, as well as high levels of test automation, there is always the potential that things fail as they progress through to, and in, production. The change failure rate looks at how many cumulative product releases fail, due to integration or other issues.

CO-ORDINATION OVERHEAD

As teams are scaled to deliver a complex or large challenges, there will need to be time added to the process to plan for success. This planning (or co-ordination overhead) is usually structured in such a way that as little is done as needed, and only in co-ordination with all parties involved in delivering a change. This is also a different activity to traditional up-front planning in traditional waterfall development, as it is not conducted by the mysterious cadre of Project Managers and PMOs, and seeks to drive out dependencies and impediments that could impede progress as opposed to committing to milestones.

MEASURES

Big room planning time - One of the key activities used to make scaled agile processes work is some form of Big Room Planning. In Big Room Planning, teams, product owners and others come together to joint plan the next "release" (usually a commercially viable product release). Given the importance of joint planning, it is usually advised that a couple of days are taken out for this activity and usually at a location away from the distractions of "the daily job" (investment in mobile phone jammers optional). This measure's purpose is to track the progress of the team commitments to meet and regularly plan together. The size of the release will dictate the number of events during the year, but the longer between sessions, the more dependencies and "wobble" will be introduced into the release. The metric can be measured in days per increment e.g. 2 days per 2 month release (or PI in SAFe parlance).

"Scrum of Scrums" time - Once a Big room event has occurred, it is expected that Scrum masters will meet on a regular basis to escalate impediments from within teams and seek to resolve issues of dependencies etc. This regular activity helps Scrum Masters and coaches look at common patterns of impediments across teams as well as devise future experiments in terms of process improvement. This measure seeks to see how regularly Scrum Masters meet. It may be that a goal is set alongside a release, or simply by measuring it, will hold Scrum Master to a regular meeting. For example, if Scrum Masters meet once a week for an hour to discuss these issues, then the measure would be 2 hours per Sprint.

Transformation Metrics

When undergoing an agile transformation, it is as important for the transformation team to measure progress as it is for teams and groups to apply measures to their own progress.

Usually transformation teams have to report into some form of management structure on the progress of the transformation, and unless the scope of the transformation is truly cross-organisational, there is usually a requirement to report progress along traditional programme or portfolio lines. Rather than looking at these traditional measures, this section only addresses potential Lean and Agile measures.

Unlike team measures that may be relevant for one, or a few sprints, and scaling measures which could apply for a larger period of time, transformation metrics are the slowest changing and longest lived of measures. That is not to say that they aren't measured on a regular basis, but that it is usually the case that they take a longer period of time to change.

MEASURES

Portfolio % delivered by agile teams – It is an exceptional organisation that commits to a wholesale change in its delivery model. Usually there are contractual and other constraints (people, location, legal, etc.) that prevent an overnight adoption of a new process. Furthermore, it is logistically challenging to ensure that people have been given sufficient investment in terms of training and support, to complete a transformation overnight. As a result of this, the Portfolio % is a measure of the amount of the change portfolio that is delivered through lean and agile methods. During early experimentation, this percentage may be very small, but over time, it can be used as a key indicator of the success of a transformation as well as helping identify where adoption may not be going well or where there may be a misalignment of people and process.

Time to market – In a large organisation there is usually a long lead time planning process, and resource strategy, that is executed prior to initiating a change within an organisation. This then leads to long running programmes of work. Lean and Agile processes encourage the creation of long lived teams working with Product Owners and product aligned backlogs of work. As a result of this the time, taken to get change released, and thus the value released, should be considerably quicker. The time to market measure looks at how long it takes for a product owner to generate an idea, and to get that change in front of customers. The measure can be applied at a low level (a simple story) through to a larger epic and can be looked in general terms across teams.

Time to terminate – Not all ideas turn out to be good ideas. In particularly, in the case of supporting digital evolution, it is important that the business can experiment with product ideas. The time to terminate metric considers how quickly a product is identified, delivered and terminated due to the fact it didn't meet the metrics specified at its inception. For example, if a business user believes a particular product can generate a reduction in customer care interactions within a 3-month period and it doesn't, the time to terminate metric looks at the overall time to mobilise a team or teams, against a backlog, reach MVP, measure progress and then terminate the product. By measuring the time to terminate, organisations are able to better judge the quality of the business cases supporting new products as well as promote a culture of experimentation.

Conclusion

A recent article [7] highlighted that dolphins who had been trained to clean their pools by giving litter to their trainers in return for a reward, were hiding and shredding larger items of litter to receive more rewards over a longer period of time. Metrics seem to arouse a great deal of passion in people, and like the case of the dolphins, there are occasions when very negative outcomes can occur. This whitepaper has presented the case that, when used wisely, metrics can help guide and improve an agile transformation and highlight where interventions are needed.

Using metrics in a repeatable way encourages coaches, and others, to constantly review the outcomes that are being achieved and to inspect and adapt the processes used to deliver change. By also regularly reviewing the metrics used to capture data, coaches can ensure that they always have the best possible view of a transformation.

The key principles we have found to be important when implementing agile metrics are:

- Use metrics as measures in complex environments to guide and inform, not just for control purposes.

- Focus on creating happy teams working to a regular cadence – regular delivery keeps teams and customers happy.

- Minimise the number of interruptions that teams suffer by aligning everyone to agile ways of working – everyone benefits from having fewer meetings.

- Reduce the number of dependencies that teams have wherever possible – give teams the ability to get to "done" by themselves.

- Apply the same standards of measurement to the actual transformation – use feedback on the process to guide the next experiment.

Finally, when 'contracting' with agile teams (be they internal or provided by 3rd parties) it is better to focus on how good partnerships can be created and value released rather than the dogmatic application of control points. Think of contracting for success, rather than what to do in the event a measure isn't met.

> "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes."

GOODHART'S LAW [4]

ABOUT ENDAVA

Endava is a privately-owned, technology company, with over 17 years of experience working with some of the world's leading Finance, Insurance, Telecommunications, Media, Technology, and Retail companies. Through our Digital Evolution, Agile Transformation and Automation solutions, we help our clients become more engaging, responsive, and efficient.

Find out more about the Agile Transformation solution we can deliver.

GET IN TOUCH

endava

REFERENCES

1. D. Vaughan, "Diane Vaughan and the normalization of deviance", https://en.wikibooks.org/wiki/Professionalism/Diane_Vaughan_and_the_normalization_of_deviance, [Accessed 05 03 2018]

2. Wikipedia, "The Observer Effect", https://en.wikipedia.org/wiki/Observer_effect_(physics), [Accessed 05 03 2018]

3. Wikipedia, "The Hawthorne Effect", https://en.wikipedia.org/wiki/Hawthorne_effect, [Accessed 05 03 2018]

4. Wikipedia, "Goodhart's Law", https://en.wikipedia.org/wiki/Goodhart%27s_law, [Accessed 05 03 2018]

5. Tuckman, "Tuckman's stage of group development", http://psycnet.apa.org/record/1965-12187-001, [Accessed 05 03 2018]

6. Wikipedia, "Andon", https://en.wikipedia.org/wiki/Andon_(manufacturing), [Accessed 05 03 2018]

7. "Dolphins are deep thinkers, https://www.theguardian.com/science/2003/jul/03/research.science, [Accessed 05 03 2018]

8. SAFe Metrics, http://www.scaledagileframework.com/metrics/, [Accessed 05 03 2018)