How Do We Know if a Scrum Master is "Good Enough" for Our Teams?

SARAH BACA, Express Scripts, a Cigna Company

In my current role at Express Scripts, I serve as one of the people leaders for our 100 Scrum Masters. Since I began this role, the biggest challenge in my mind was: How do we know our Scrum Masters are doing a good job? And what does it mean to do a good job in our highly contextual roles as Scrum Masters?

In this experience report, I will share how we are trying to answer this question, and how it is evolving as we learn more.

1. INTRODUCTION

In my role as a people leader, I want our Scrum Masters to be as effective as they can be so that our teams can function well and create software that helps our patients. But how do we know if a Scrum Master is doing a good job? What does that mean in our roles when so much depends on the dynamics and personalities of team members, the leaders in other organizations, and the established systems around the teams?

2. BACKGROUND

At Express Scripts we have about 80 Scrum Masters and 20 agile coaches. I'm one of the four people leaders who lead the Scrum Masters. When I first came into this role I questioned how I could bring value to my team of about 20 people. A manager's role is rarely defined in most agile literature, and a Scrum Master manager most definitely isn't. I've always been passionate about motivating and supporting people. So how could I translate that to my new role? That has been an interesting and sometimes painful lesson that I continue to learn.

3. YOUR STORY

When I was a Scrum Master, an important part of my job was understanding where my team was at, the next step to coach towards, and how to keep track of that progress. If I understand their pain, it helps me develop empathy so I can work with them as we grow together. In my role as a Scrum Master people leader, it's my job to support and help other Scrum Masters and their teams through that same journey while applying those skills at a higher level, to the leadership of other organizations.

I work with three other Scrum Master people leaders, each of us having 20-30 contractors and employees on our teams. The four of us work together closely, bouncing ideas off each other and challenging each other. Any good ideas in this report came as a result of our collaboration.

3.1 Challenges

When I first started in this role, I thought it was my job to measure if Scrum Masters were competent enough to do their jobs. I still believe it's my responsibility, if it becomes clear that someone isn't serving their team effectively, to work with the team to solve that problem. That sometimes includes ending a Scrum Master's time with us, and the difficult conversations that are part of that. However, I've learned now that if I can support, encourage, and focus on creating a system where Scrum Masters can be successful, I still grow to understand their competency level. And by approaching it that way, I can help them grow, whereas if I am solely judging them, I've closed off any trust that might cause them to be open to a coaching relationship with me.

A tangential topic to Scrum Master competency is team maturity. Our company has gone through several cycles of trying to measure team maturity. As I was learning that focusing on judging Scrum Master competency wasn't helpful, I also learned that measuring team maturity wasn't helpful. This paper goes into

more detail on how I grew (and continue to grow) a deeper understanding of what is helpful to focus on, instead of spending time and energy measuring everyone's maturity and competency. We've learned ways outside of a maturity assessment to get quantitative data (such as cycle time and release frequency) and combine that with qualitative data (conversations and stories) to figure out how best to help our teams and our patients. These methods have proven to be more informative and helpful than others.

3.2   Team-Based Assessments

We have tried several different ways to measure team maturity and Scrum Master competency.

When our transformation first began, a large consultancy was hired to lead it and train the teams. They created a survey for the team to take together that they called the Agile Maturity Assessment (AMA). The entire team was expected to take the survey, it took an hour or two to complete, and they were expected to take it every two weeks.

The AMA was in Excel and had a scale for each item.

Legend	Legend	Definition
N/A	Not Applicable	This sub-category is either not applicable to the team, not quantitatively or qualitatively viable, or cannot be started due to constraints outside of the team's control
1	Challenged	Does not meet the assessment's criteria, the team is unable to improve it due to impediments (of which can be remedied if action is taken on them), and/or they have a differing view on whether the sub-category is relevant to their team.
2	Needs Improvement	Does not meet the assessment's criteria, and nothing is being done to currently improve it
3	Improving	Does not meet the assessment's criteria, and the team agrees that this area needs to be improved and is making efforts to improve it
4	Stable	Does not meet the assessment's criteria, but the team agrees that it is in a state that works for them. The team is open to working to the exact criteria if it's not too disruptive.
5	Ideal	Meets the assessment's criteria and no work needs to be done in this sub-category.

Table 1. Agile Maturity Assessment Scale

There were both foundational and advanced categories for the questions. The team would go through each and rate themselves for each item and Excel would average the totals for each category. Items included topics like: "at least two sprints worth of work is in the backlog" and "team is not creating separate stories for dev and QA in their backlog."

Some benefits of this survey were that it made it possible to measure maturity over time. It also made it clear what "good" looked like, so the team knew the desired state, based on their interpretation of the category. The team was able to have conversations while going through the AMA about how to get their score higher and what behaviors might be necessary to get the team to a level five.

There were several drawbacks to the AMA. When team members rated themselves, they might not feel safe to be honest if there were negative consequences for lower numbers or for not moving to higher numbers quickly. The team members also might not know what good really looked like and might interpret the sub-category differently than an experienced coach would. It also took several hours of time from the entire team, making it a very expensive assessment.

Back when I was a scrum master, I would sit with the members of my two teams and take this assessment. Often we would cheat and I would do it with the development leads representing the team because the team felt like doing the assessment was a waste of time. Of course, the survey couldn't cover everything, so there would be areas where I knew the team needed to grow, even if we looked like we were high-performing according to the scale. The team would grumble and complain about all the time they lost when they did take this assessment, and I got the impression that they felt like the agile transformation office didn't care about them—only what scores they got. Our leadership did care, but all they heard was that our leadership wanted them to take this survey.

Our next iteration of the assessment was called the Agile Health Assessment. These assessments were taken individually and there were only ten questions, so it took much less time to complete. Also, they completed this assessment individually and when it was convenient for them, versus a long meeting.

This is the survey as our team members saw it:

> To continue in our agile transformation, we need your thoughts and feedback. The Agility Health Assessment (AHA) does not rate your skill level or team competency. Instead, it serves as a forum to express how you feel about our progress in our agile journey. The data from the AHA is aggregated to show how the team members feel about team agility.
>
> We know your time is valuable, so the survey is only 15 questions long and it is supposed to be taken individually. We take personal privacy seriously and have done our best to aggregate the data. Thank you for taking the time to respond openly and honestly.
>
> Let's get started...
> Please answer these questions from your experience and perspective in the Business Portfolio and Technical Domain. You may be asked to complete this survey more than once, to keep these combinations separated.
>
> Options are: Always; Most of the time; Sometimes; Rarely; Never
>
> 1) We understand clearly what we will work on next and why:
> 2) We recognize good work and celebrate success:
> 3) We actively experiment with our product, process and craftmanship to improve our solutions and our team:
> 4) All team members express their opinions/concerns:
> 5) We measure the impact of WIP limits on our cycle time and adjust accordingly:
> 6) We adjust and abandon opportunities and solutions upon feedback:
> 7) We trust each other:
> 8) We make everyone around us awesome:
> 9) We frequently deliver value to our end users:
> 10) We are motivated by our work and have the opportunity to improve our craftsmanship:
> 11) We reflect on how to become more effective then tune and adjust our behavior accordingly:
> 12) We are not afraid to fail and we learn from our failures:
> 13) We can change our code easily and safely when a new change is requested:
> 14) We can describe how our work improves the lives of ESI patients and clients:
> 15) We make informed decisions upon listening to different opinions than our own:
>
> We Value your feedback! Please share your comments below...

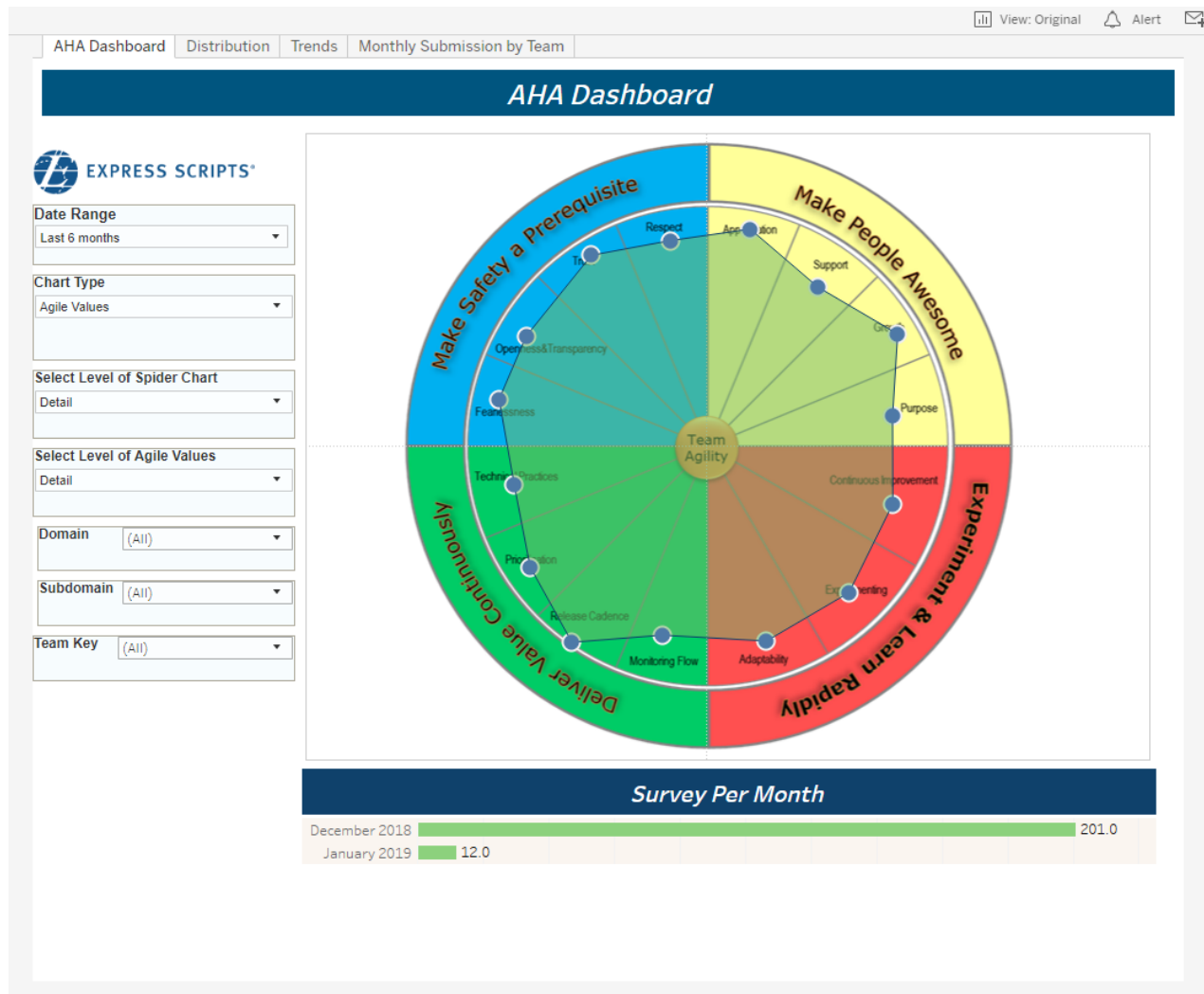The results were exported to a spider graph for easy viewing.

Figure 1. Agile Health Assessment Spidergraph

The challenge with this survey was that the team was still assessing themselves, and many of the team members didn't know what "good" looked like. We saw teams rating themselves high on categories like collaboration, when Scrum Masters and coaches were seeing siloed behavior.

The teams took this survey after I had moved to a leadership role, so I didn't take it with any teams. But I did talk to my scrum masters about their teams taking it, and even though it was quick, they had a really hard time getting their teams to take it. The best solution we found was getting teams to take it during their retrospectives, which of course meant they were losing time reflecting on what actually worked for them. Because it didn't give an accurate representation of how the teams were doing, the scrum masters were frustrated. They still had no data on how they could help the teams.

Culturally, we speculated team members also felt like if they weren't scored highly enough, there might be negative consequences. People were motivated to artificially rate themselves higher, which meant we didn't know where they needed help. The survey showed teams were very mature when we knew they weren't. We needed a better solution.

3.3   Assessment Based on the Twelve Agile Principles

Our next iteration started out as an assessment created by my peers and me, the Scrum Master people leaders (SMPL), based on the twelve agile principles. We called it the Twelve Agile Principles Assessment (TAP Assessment). We created guidelines and asked Scrum Masters to measure their teams in green, yellow, red, with an arrow trending up (getting better), sideways (staying the same), or down (getting worse). By having the Scrum Masters fill out the TAP Tool, we could determine the understanding level of the Scrum Masters—

like what kind of evidence they rated green versus yellow. We also could discuss with the Scrum Masters how they planned to teach and coach their team to get to green. This exposed information about their skillset while also giving us a way to measure the teams.

When we started using the TAP Assessment, we thought that we had finally found a way to measure team maturity. Our Scrum Masters knew what good looked like so they could assess their teams and if they didn't know, by going over it with us, we could tell them what good looked like.

When we started reviewing the TAP Assessments we were happy to see that we were right. Some of our Scrum Masters who knew what good looked like could tell us where their teams were and how they were coaching their teams. Some Scrum Masters thought their teams were mature but upon further conversation it was clear that they didn't understand what maturity meant for that level. We could teach them what behavior was expected for that category and they could learn. Success!

But we also saw some behaviors that concerned us. A lot of our contractors said their teams were all green, even though the scrum masters knew they weren't. We people leaders knew from previous conversations with these scrum masters that these teams weren't green, but it was clear the Scrum Masters were afraid of the consequences of having a team that was struggling. Some Scrum Masters were obviously afraid that if their team wasn't all green, they would lose their jobs.

We started to question what behaviors we were trying to encourage and how what we were measuring encouraged them. We began to realize that our roles were more effective when we "improve the system, not the rules, or the people … a self-organizing system of competence is the only maturity level you will ever need." (Appelo, p. 246-7).

Also, during the time our Scrum Masters were completing these assessments and we were reviewing them, our leadership was starting to understand more about what the Scrum Master role. They started to recognize that it's difficult to define a fully matured state for an agile transformation—we can always get better. While we are all on a journey, there's no clear mountaintop we are traveling to. That mountaintop is ahead of us, but it changes so often and evolves so much, that we will never really arrive. Where we are tomorrow is relative to where we are today, based on the experience of the team, the stakeholders, and the Scrum Masters.

3.4 Assessment Results

After many meetings to discuss how we could use the TAP Assessment to measure our Scrum Masters and our teams, we realized that we were spinning in circles. Each time we dug deeper and asked more questions, we ended up in the same place—trying to measure our Scrum Masters and our teams like this didn't make sense.

A better way to understand our success isn't a red, yellow, and green spreadsheet; it's stories told with data, that are told by the people who lived them.

We realized that measuring maturity and fighting to get to the next color wouldn't work. We were fighting against the current reality, rather than creating something that inspired people to change. "You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete." (Laloux, p. 1). We instead needed to gather several metrics that could help us understand the state of the system, and partner with others outside of our organization to shift perspectives.

Since a maturity assessment isn't going to help us, we gather metrics that give us data that will actually help our patients. These are the metrics/graphs we are currently experimenting with:

1. User story cycle time
2. Story point cycle time
3. Acceptance rate (measuring completed vs. committed work)
4. Average story size in points
5. Release frequency
6. Cumulative Flow Diagram
7. Ratio of time spent on new features vs. maintenance vs. defects
8. Active production defects

There is no one metric that can tell us if we're "mature" or not. It's too contextual and too complex to measure that way.

There was value in the conversations we had with our Scrum Masters using the TAP Assessment, so we changed it to be a tool we use to guide our one-on-one conversations with Scrum Masters. The value of those

conversations, and the growing shared understanding, is huge. Focusing on conversations also allows us to spend the time and energy that we would be spending on gathering data and measuring in other ways—like learning what our customers want and focusing on outcomes that they want. Maturity doesn't matter at all if we're not creating outcomes that our customers want.

In addition to the TAP Tool, we have a few other ways that we are gathering subjective qualitative and quantitative data.

3.5   Coaching Plans

One of those ways is a coaching plan. As we review the TAP Tool with the Scrum Masters, each Scrum Master updates a coaching plan. They maintain this plan in between TAP Tool reviews with their people leader, and use it to track how they are helping our teams over time.

The coaching plan has a column for:
- What behavior the SM is seeing (evidence)
- What beliefs the team has that is causing this behavior (inference)
- What behavior the SM thinks should be happening (to check knowledge)
- What the SM will be doing/coaching toward to get closer to what they think should be happening (impact)

3.6   Servant Leadership Survey

We combine this with another tool that we started using, a survey that goes out to the team members. We call this survey the Servant Leadership Survey and send it out quarterly.

These are the questions on the Servant Leadership Survey. All response options are: Strongly agree, Agree, Neutral, Disagree, and Strongly Disagree.
1. My Scrum Master promotes a culture where my voice is heard and it's okay to disagree.
2. Team members feel safe to speak up during regular team check-ins (stand up, etc.) without prompting.
3. The entire team has the opportunity to become familiar with the work before they start working on it.
4. I would recommend this Scrum Master to another team (we promise we won't move them!).
5. My Scrum Master encourages me and my team to grow and continually improve.
6. My Scrum Master raises awareness and facilitates resolution of impediments, encouraging the team to take ownership as appropriate.
7. My Scrum Master encourages collaboration across roles and across different teams.
8. My team has a regularly-updated working agreement that is owned by the team.
9. My Scrum Master protects the team while inviting them to advocate for themselves.
10. My Scrum Master makes suggestions rather than telling the team what to do.

We know that this survey has similar problems to the other surveys—different understanding of what each phrase means, a possible fear of recourse if they score negatively. We decided that risk was worth it, as this is a simplified anonymous 360 for our Scrum Masters, and we want that feedback even if it's not perfect.

Understanding the results of the servant leadership survey isn't a simple process for the Scrum Master people leaders (SMPLs). Sometimes a team will love their Scrum Master because the Scrum Master doesn't challenge them. Sometimes a team will dislike their Scrum Master because the Scrum Master has just started challenging them, and they're not used to it. We use this as another data point during our one-on-one conversations with our Scrum Masters, so we can help them grow and learn.

3.7   Communities of Practice

We also started several Scrum-Master-led communities of practice. These are small groups (called guilds) that the Scrum Masters use to learn from each other. Currently we have a coaching guild; a "stump the coach" guild, where the Scrum Masters interview coaches or another Scrum Master; a Kanban guild; a technology guild; a clean language and systemic modeling guild; and more that pop up when people become interested in a topic, and dissolve when interest in that topic wanes.

All of these methods give us more information so we can support our Scrum Masters and teams. This creates a better organizational culture and it also helps our teams so we can help our patients.

4. WHAT I LEARNED

While I was working on understanding how to measure the teams' maturity and scrum master competency, some of our coaches started teaching a leadership course. Part of this course included a coaching session with a co-active coach, Jennifer Davis. She asked me to coach me in a short coaching demo at the front of the room and I agreed. During this session, she kept digging into why I felt that I needed to measure my team. Why did I feel like it was so important to assess them? What was I hoping to gain by measuring where we were in our journey? What exactly was I trying to accomplish? It was then that I realized that what I was trying to prove wasn't really about my team or our agile journey. What I was trying to prove was that I had value as a leader and that I knew what I was doing. By trying to judge others, I was asserting my superiority and establishing my worth.

If I could release that need to prove that I was a valuable person and that I was somehow better than everyone else, I could spend that time and energy on supporting a system that serves our Scrum Masters. Instead of judging team maturity, I could spend my time helping Scrum Masters decide on next steps to help the team. When I support instead of judge, I'm able to be more effective because my relationships with those around me are stronger. I'm not focused on proving how good I am—I'm focused on helping those around me to become better.

I hadn't done that yet because letting down the protection of judging others left me vulnerable for others to see that I might not know what I was doing, and I might not actually bring value. That was scary to me, so I had worked to design elaborate ways of judging others to protect myself. Those protections had left me much less effective and had actually hurt the people I was trying to help. By having the courage to release the need to judge others, I could become the leader I had wanted to pretend I was.

Now instead of spending time assessing and measuring Scrum Masters, I work with leaders in other groups to try experiments that might make our Scrum Master's jobs easier. We ask the leaders in those groups and asking "What challenges are you experiencing? How can we help?" The coaches and Scrum Masters work together to create plans for solving those problems. I meet with the Scrum Masters on my team and find out what challenges they are experiencing and ideas they have for how I can help. We're starting to learn how work gets to the teams, to reduce the overall cycle time and make sure that the work the team receives actually helps our customers and patients. We're questioning if we can reduce dependencies between the teams. By focusing on helping the system, I'm able to bring value to our entire organization.

5. ACKNOWLEDGEMENTS

REFERENCES

Appelo, Jurgen. Management 3.0. Pearson Education, 2011.
Laloux, Frederic. Reinventing Organizations. Nelson Parker, 2014.
Jennifer Davis Coaching, https://www.jenniferdaviscoaching.com/