Harri Besceli - Notes on Metrics

Goals and Metrics

Goals are explicit aims that we think will help us achieve our values. For examples, my goal is to run a marathon, and I have chosen this because I value being healthy, pushing myself and sweating.

Making goals explicit makes it possible to break it down into sub-goals, (which can then be broken down into actionable steps). By making goals explicit you can more easily identify actions that will progress you towards your goal, making them easier to achieve.

A metric is a quantitative measure that determines a success condition for a goal. For example, I value my happiness. I may set the goal ‘Rate my mood 7/10 or higher on average over a period of a week, using experience sampling’. The metric is the average mood rating over a week.

Quantitative measures serve to make it easier to track whether you are making progress towards a goal. If your weekly average rating has increased from 5/10 to 6/10, you know you are making progress. This feedback helps you pick the right tactics for achieving the goal. If your weekly average rating hasn’t increased after doing positive affirmations for 2 weeks, it may be time to try a new tactic. Metrics are particularly valuable for long-term goals where the actions which will get you there aren’t obvious.

Lead measures, Lag measures and theory of change

A metric for a goal is a lag measure, whereas a metric for a subgoal is a lead measure. A subgoal for ‘Rate my mood 7/10 or higher on average over a period of a week, using experience sampling’ could be ‘Meditate once a day for one month’. The lead measure is number of days meditated in the month.

Good lead measures track processes or events that cause an increase in the lag measure. We can call our model of the relationship between a lead measure and a lag measure a *theory of change*.

For example, my theory of change may be, regular meditation causes me to relax, and relaxing improves my mood which is a constituent of happiness. So the number of days on which I meditate (lead measure) increases, my weekly average mood rating should increase (lag measure).

Lead measure are easier to influence directly but are often less correlated with the goal than lag measures. With an accurate theory of change, we can influence lead measures, which will influence our lag measure and achieve our goals.

Qualities of Good Metrics

1. **Informativeness:** How much does the metric tell us whether we have achieved our goal? How well do mood ratings tell me how happy I am?
    a. **Accuracy:** How closely does the metric correlate with what you value/ your goal?
    b. **Resolution:** How fine-grained are the measurements? For example mood ratings out of 100 have a higher resolution than mood ratings out of 10.
    c. **Precision:** How closely correlated are the measurements? Given the same inputs, how much would the scoreboard (the metric) vary? For example, how much would my ratings of my mood vary, when I am in the same mood?
2. **Usefulness:** How easy is the metric to use?
    a. **Tractability:** How easy is it to calculate the total score?
    b. **Feedback:** How long is the delay between interventions aimed at changing the score and the score itself? How long does it take for meditation to have an impact on my mood ratings? Quicker feedback means the theory of change can be tested and improved more quickly.
    c. **Predictable:** How well can you predict the score? How well can I predict my future mood ratings with other information. Scores are predictable when you have a good theory of change.

Goodhart's law: Goodhart's law, **'**When a measure is used as a target it ceases to be a good measure' is a potential failure mode of using metrics. When something which is usually correlated with value is chosen as an explicit goal, the correlation between the goal and what is valuable weakens.

For example, number of hours spent in the office (the lead measure) is usually a good measure of doing productive work (the goal), measured by number of pomodoros on priority tasks. Increasing this may have no effect, or a negative effect on doing productive work.

This is because the actions are based on an inaccurate theory of change. Spending time in the office may enable productive work, but the amount of productive work may be limited by energy levels. So going to the office doesn't cause an increase in productive work.

Goodharting can happen to both lead and lag measures, for example, the lag measure of number of pomodoros on priority tasks may also not result in an increase in productive work.

Countermetrics can help avoid ‘Goodharting’. For example, an additional lead measure may be ‘number of hours exercised’, based on the theory of change that exercise increases energy levels.

More information on Goodhart’s Law here and here.

Examples of Metrics

Value	**Lead Measure**	**Lag measure**	**Goal**	**Theory of change**	**Goodhart/ potential problems**	**Countermeasure/ change**
Increase stamina	Number of time put on gym clothes.	Number of times gone to the gym.	Go to the gym 5 times a month for the next 3 months.	Putting on my gym clothes will cause me to go to the gym because it reduces the barrier to entry.	**I** put on my gym clothes but don’t go to the gym.	If I put clothes on but don’t go the the gym more than 3 times, change strategy.
Learn German	Number of pomodoros spent studying german.	Number of conversations (of 10m+) had in German, where I didn’t have to revert to speaking English.	Have 10 conversations in German.	Studying german will improve my conversation ability in German.	I may only read German whilst studying, which doesn’t improve my conversational ability.	