What We Learned from Four Years of Sciencing the Crap Out of DevOps

Nicole Forsgren, PhD
CEO and Chief Scientist
DevOps Research and Assessment (DORA)
Outline

How to make your data suck less
• Writing good survey questions
• Making sure the survey questions are good - with SCIENCE
• (These methods apply to your system and log data)

What we found... that we did (AND didn’t) expect
Things about Continuous Delivery
Things about Management

@nicolefv
Not all data is created equal

Who here thinks surveys are sh*t?
Not all data is created equal

Who here thinks surveys are sh*t?

Who here LOVES the data from their log files?
What is a Latent Construct?
We use **PSYCHOMETRICS**

to make our survey data good*

*or give us a reasonable assurance that it’s telling us what we think it’s telling us (& some of this can also apply to your log data)
Psychometrics includes:

Construct creation (manual)
- When possible: use previously validated constructs
- Based on definitions and theory, carefully and precisely worded, card sorting task, pilot tested

Construct evaluation (statistics)
- Establishing Validity: discriminant and convergent
- Establishing Reliability
Psychometrics Writing Example: Culture

• Does it matter to our study?
  • More than just intuition?
• What KIND of culture?
  • National identity and norms
  • Adaptive culture
  • Value learning (2014 study)
  • Value information flow and trust (2014 and 2015 studies -- Westrum culture)
Psychometrics Writing Example: Culture

- Does it matter to our study?
  - More than just intuition?
- What KIND of culture?
  - National identity and norms
  - Adaptive culture
  - Value learning (2014 study)
  - Value information flow and trust (2014 and 2015 studies -- Westrum culture)
Try writing items yourself!

Use strong statements with clear language.

<table>
<thead>
<tr>
<th>Pathological</th>
<th>Bureaucratic</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Power-oriented</em></td>
<td><em>Rule-oriented</em></td>
<td><em>Performance-oriented</em></td>
</tr>
<tr>
<td>Low cooperation</td>
<td>Modest cooperation</td>
<td>High cooperation</td>
</tr>
<tr>
<td>Messengers shot</td>
<td>Messengers neglected</td>
<td>Messengers trained</td>
</tr>
<tr>
<td>Responsibilities shirked</td>
<td>Narrow responsibilities</td>
<td>Risks are shared</td>
</tr>
<tr>
<td>Bridging discouraged</td>
<td>Bridging tolerated</td>
<td>Bridging encouraged</td>
</tr>
<tr>
<td>Failure leads to scapegoating</td>
<td>Failure leads to justice</td>
<td>Failure leads to inquiry</td>
</tr>
<tr>
<td>Novelty crushed</td>
<td>Novelty leads to problems</td>
<td>Novelty implemented</td>
</tr>
</tbody>
</table>

Westrum Culture Items

- On my team, information is actively sought.
- On my team, failures are learning opportunities, and messengers of them are not punished.
- On my team, responsibilities are shared.
- On my team, cross-functional collaboration is encouraged and rewarded.
- On my team, failure causes inquiry.
- On my team, new ideas are welcomed.
Psychometrics Analysis
Example: Notification of Failure

At my organization...

- We are primarily notified of failures by reports from customers.
- We are primarily notified of failures by the NOC.
- We get failure alerts from logging and monitoring systems.
- We monitor system health based on threshold warnings (ex. CPU exceeds 100%).
- We monitor system health based on rate-of-change warnings (ex. CPU usage has increased by 25% over the last 10 minutes).
At my organization...

- We are primarily notified of failures by reports from customers.
- We are primarily notified of failures by the NOC.
- We get failure alerts from logging and monitoring systems.
- We monitor system health based on threshold warnings (ex. CPU exceeds 100%).
- We monitor system health based on rate-of-change warnings (ex. CPU usage has increased by 25% over the last 10 minutes).
More data tests!

Plus, we test to make sure the survey doesn’t have other problems.

- Common method variance (CMV) (aka CMB for Bias)
- Early vs. late responders
- Survey drop-off rates and bias
Okay **NOW** we can look at the data and how it relates to each other
A note about analysis methods

One of three conditions must be met:

1. Longitudinal (no, this is cross-sectional)
2. Randomized, experimental design (no, this is a non-experimental)
3. Theory-based design

When this condition was not met, only correlations were tested and reported
KEY FINDING: IT Performance and its behavior

A combination of \textit{throughput} and \textit{stability}

- lead time for changes
- release frequency
- time to restore service
- change fail rate

KEY FINDING: IT performance matters!

“Firms with high-performing IT organizations were twice as likely to exceed their profitability, market share and productivity goals.”

IT Performance is predictive of organizational performance.


### 2016 IT Performance by Cluster

<table>
<thead>
<tr>
<th></th>
<th>High IT Performers</th>
<th>Medium IT Performers</th>
<th>Low IT Performers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deployment frequency</strong></td>
<td>On demand (multiple deploys per day)</td>
<td>Between once per week and once per month</td>
<td>Between once per month and once every 6 months</td>
</tr>
<tr>
<td><strong>Lead time for changes</strong></td>
<td>Less than one hour</td>
<td>Between one week and one month</td>
<td>Between one month and 6 months</td>
</tr>
<tr>
<td><strong>Mean time to recover (MTTR)</strong></td>
<td>Less than one hour</td>
<td>Less than one day</td>
<td>Less than one day*</td>
</tr>
<tr>
<td><strong>Change failure rate</strong></td>
<td>0-15%</td>
<td>31-45%</td>
<td>16-30%</td>
</tr>
</tbody>
</table>

* Low performers were lower on average (at a statistically significant level), but had the same median as the medium performers.
Together, the factors on the left model "Continuous Delivery", which leads to...

- Effective test data management
- Comprehensive, fast and reliable test and deployment automation
- Trunk-based development and continuous integration
- Application code and app and system configuration all in version control
- Incorporating security — and security teams — into the delivery process

Less rework
- Lower levels of deployment pain
- Higher levels of IT performance (higher throughput and stability)
- Lower change fail rates

Identifying strongly with the organization you work for
- Higher levels of org performance (productivity, market share, profitability)
some surprises
Which of these measure effective test practices?

Developers primarily create & maintain acceptance tests
QA primarily create & maintain acceptance tests
Primarily created & maintained by outsourced party
When automated tests pass, I’m confident the software is releasable
Test failures are likely to indicate a real defect
It’s easy for developers to fix acceptance tests
Developers share a common pool of test servers to reproduce failures
Developers create on demand test environments
Developers use their own dev environments to reproduce failures
Which of these measure effective test practices?

- Developers primarily create & maintain acceptance tests
- QA primarily create & maintain acceptance tests
- Primarily created & maintained by outsourced party
- When automated tests pass, I’m confident the software is releasable
- Test failures are likely to indicate a real defect
- It’s easy for developers to fix acceptance tests
- Developers share a common pool of test servers to reproduce failures
- Developers create on demand test environments
- Developers use their own dev environments to reproduce failures
Key Finding: Unplanned work

High performers spend 20% more time on new work than low performers, and 22% less time on unplanned work and rework.
Surprises with Culture

• We wanted to add additional measures of culture
  • Google study
  • Identity
  • Retain Westrum culture
Identity & Google items

- I am glad I chose to work for this organization rather than another company.
- I talk of this organization to my friends as a great company to work for.
- I am willing to put in a great deal of effort beyond what is normally expected to help my organization to be successful.
- I find that my values and my organization's values are very similar.
- In general, the people employed by my organization are working toward the same goal.
- I feel that my organization cares about me.


@nicolefv
Identity & Google items

• I am glad I chose to work for this organization rather than another company.
• I talk of this organization to my friends as a great company to work for.
• I am willing to put in a great deal of effort beyond what is normally expected to help my organization to be successful.
• I find that my values and my organization's values are very similar.
• In general, the people employed by my organization are working toward the same goal.
• I feel that my organization cares about me.

Now for management stuff

We all know managing WIP is important, right?
Now for management stuff

We all know managing WIP is important, right? Correlation between WIP and ITPerf is negligible
Now for management stuff

We all know managing WIP is important, right? Correlation between WIP and ITPerf is negligible. What’s going on?
Lean management SEM

Effective WIP limits that drive process improvement

Use of visual displays to monitor quality, productivity and work in process

A generative, performance-oriented culture (per Westrum's model)

Higher levels of IT performance (higher throughput and stability)

Higher levels of org performance (productivity, market share, profitability)

Together, the factors on the left model "Lean Management", which leads to...

Use of app perf and infra monitoring tools to make business decisions

Lower levels of burnout
Also lean product management

- Gathering, broadcasting and implementing customer feedback
- Splitting work into small batches and making the flow of work through the delivery process visible
- Together, the factors on the left model "Lean Product Management", which leads to...
  - Higher levels of IT performance (higher throughput and stability)
  - Lower levels of deployment pain
  - A generative, performance-oriented culture (per Westrum's model)
  - Identifying strongly with the organization you work for
  - Higher levels of org performance (productivity, market share, profitability)
Conclusions

• Even if you think it’s obvious, TEST WITH DATA.
  • (if the results don’t surprise you, you’re doing it wrong)
  • (if you don’t also confirm some things you expected, you’re doing it wrong)
• We CAN have it all, or at least throughput AND stability.
• IT matters (but you have to do it right)
• DevOps culture & practices have a measurable impact on IT & org perf
For more science-ing...

Sign up for our ROI whitepaper & get peer-reviewed research

devops-research.com

@nicolefv
Thank you
devops-research.com