

## 02 | 效能度量：效果不好甚至有副作用，怎么回事？

2019-08-26 葛俊

研发效率破局之道

[进入课程 >](#)



讲述：葛俊

时长 14:26 大小 13.23M



你好，我是葛俊。今天，我来和你聊聊研发效能的度量。

在和技术管理者，尤其是高层管理者聊起研发效能的时候，常常提起效能的度量这个话题。管理学大师彼得·德鲁克（Peter Drucker）曾经说过，“一个事务，你如果无法度量它，就无法管理它”（If you can't measure it, you can't manage it）。要想提高研发效能，自然要首先解决效能的度量的问题。

在软件研发过程中，数据驱动的手段被大量采用，比如说通过使用漏斗指标和 A/B 测试，用数据来指导产品的方向。按理说，软件开发行业是数据驱动的高手，用数据驱动来理解研发效能应该早被研究透了啊。

但事实上，效能的度量却是一个出了名的难题，至今没有哪个公司敢号称已经找到了效能度量的完美答案。不仅如此，绝大部分软件公司在使用研发效能度量这个工具时，不但没有起到正向作用，还伤害了产出和团队氛围。

所以，在今天这篇文章中，我就与你一起看看研发效能度量到底是什么、度量使用中的常见错误，以及困难的原因。

## 研发效能度量的定义和作用

首先，我来介绍一下效能度量的定义和作用。

研发效能的度量代表一组可量化的数据或参数，用来跟踪和评估开发过程的“健康”状况。换句话说，**研发效能的度量，从应用程序开发生命周期中获取数据，并使用这些数据来衡量软件开发人员的工作效率。**我们希望通过这样的度量，能够根据客观的数据而不是个人的主观意见去决策，从而实现以下几点：

1. **跟踪团队的表现、提高团队的绩效。**通过确定研发效率指标，公司可以明确团队和成员的工作预期，从而使得开发人员能够目标性更清晰地投入研发。同时，这些生产指标可以作为“晴雨表”，帮助团队定位影响工作效率的不良因素，从而帮助团队消除这些因素，最终达到团队更快乐、绩效更高的目的。
2. **提高项目计划的精确度。**团队负责人可以通过度量来估算一个需求端到端的成本，包括收集成本、设计系统成本、开发测试成本，以及运维成本等，来了解每项活动在项目总成本中的占比，从而更好地确定这些活动的优先级。同时，我们可以了解哪些步骤有较大风险和不确定性，从而提高预测项目进展的精准度。
3. **了解流程是否高效，寻找需要改进的关键领域。**我们可以衡量进行每项研发活动所需的时间，并估算其对质量和生产效率的影响，然后比较成本和收益，最终确定哪些步骤是高效的，以及哪些步骤是需要改善的。我们还可以对不同的实践进行 A/B 测试，以此来选择更好的方法。

提高团队绩效、提高计划精确度，以及寻找关键的待改进领域，这三个因素的结合有助于简化工作流程，发现瓶颈，帮助团队持续改善现有产品的生命周期，从而更高效地生产出质量更好的产品。

## 效能度量的错误案例

因为上述作用，绝大多数公司都非常重视研发效能的度量，并且基本上所有公司都或多或少地采用了度量。但遗憾的是，能够将其用好的公司少之又少，而且我见到的公司中还存在大量误用的情况。

接下来，我与你分享一些真实案例，我们一起来看看它们错在哪里，以及可以如何调整。

## 案例 1：全公司范围内推行一套效能度量指标

某大型公司为了提高效能，专门成立了一个不小的团队，通过详细调研，制定了一组研发效能度量指标，并引入和开发了相应的工具。客观来讲，这些度量方式的质量很高。

准备就绪后，这个公司找了几个团队做试点，把这些效能度量数据作为参考提供给他们使用。三四个月下来，试点效果不错，这几个团队的产出速度有了一定的提高。于是，公司高层决定在全公司范围内推行这套方案。

但，随着大范围的推广，一些团队发现，它们的研发流程跟试点团队差别较大，需要花费相当多的时间去收集度量数据，受益却不明显，甚至是得不偿失。

更极端的情况是，有一个团队的开发环境和这套效能度量工具的环境差别实在太大，所以不得不专门为这个工具部署了一个环境。然后每个周六，他们就要到这个环境里，为度量系统提供数据，还常常被迫提供一些“假数据”，以满足度量系统的需求。

因此，这些团队觉得这套方案并不适合自己，公司内逐步出现了些许反对声音。但是，公司推进这套流程的态度非常强硬，为了顺利推进，还把这套流程与团队绩效绑定到了一起。一旦某个团队的指标未能达到要求，就直接实行绩效扣分。由此，效能度量的负面作用逐渐凸显。大量团队为了好的绩效，而被迫去玩数字游戏。

全公司范围推行这套方案的一年多，可以说是怨声载道。明明是一套高质量的度量系统，却阻碍了团队业务的发展，并严重伤害了员工的积极性。直至最后这个实际情况被反馈到一位高管那里，这套度量系统才被完全取缔。

## 案例 2：一个中型公司推行质量方面的指标

一个 500 人左右的公司，公司的组织架构按照职能进行水平划分，并没有进行矩阵式管理。为了提高公司的产品质量，QA 团队提出了一组软件质量方面的度量指标，包括 Bug

严重性定义，每一个发布里不同级别的 Bug 的未解决率，以及上线的一些质量流程。QA 团队得到了公司高层的支持，强制推行这些指标。

这些指标针对的主要是开发活动，但开发团队却认为，这些度量用处不大。他们认为，公司研发过程中的真正问题在于产品定义变化过快，常常是在冲刺开始后还不断更改需求。这就导致开发和测试团队即使拼命加班还常常错过发布日期，产品的 Bug 也不少。所以，开发团队提出，真正应该度量的是产品需求的稳定性。但是，因为公司高层认为需求变化是为了满足用户需求，所以没有批准开发团队的这个要求。

结果是，开发团队觉得公司效能度量只是用来束缚他们的，根本不能帮助他们工作，于是工作积极性下降，离职率上升。同时，产品质量也并没有因为这些度量指标的推行而提高。直到我写下这篇文章的时候，这个公司还是这个情况。

### **案例 3：某创业公司聚焦度量开发、测试、上线准确度等指标**

一个由 10 多个人组成的，拿到了 Pre-A 投资的创业公司，正处在研发产品寻找市场吻合度的阶段。掌权的 CEO 和产品副总坚信数据驱动，于是对研发流程定义了严格的度量和指标，比如 App 上线周期准时率、Bug 的关闭速度、性能参数，等等。

整个团队很专业，也很敬业。他们花了很多精力去严格完成这些度量指标，做到了绝大部分情况下准时、高质量地上线产品。但是，CEO 和产品副总并不是产品专家，只关注了开发过程中的数据，却没有收集其他步骤的数据去快速试错和寻找产品方向。

结果就是，虽然产品的每一个发布都很准时，质量也非常高，但因为公司在寻找市场需求吻合度方面动作迟缓，导致用户增长缓慢。因此，一年半以后，资金耗尽，投资人失去信心，公司倒闭。

这 3 个案例只是不同规模公司的几个典型场景，类似的失败案例数不胜数。那么，上面这些案例的问题出在什么地方？效能度量为什么这么难？

### **效能度量被大量误用，问题究竟出在哪儿？**

**研发效能难以度量的最根本原因在于**，软件开发工作是一项创造性很强的知识性工作，非常复杂且伴随有大量不确定因素。

比如，软件产品的需求变化很快，需求文档的更新常常滞后于工程实现，甚至有的敏捷方法论提倡完全抛弃需求文档。

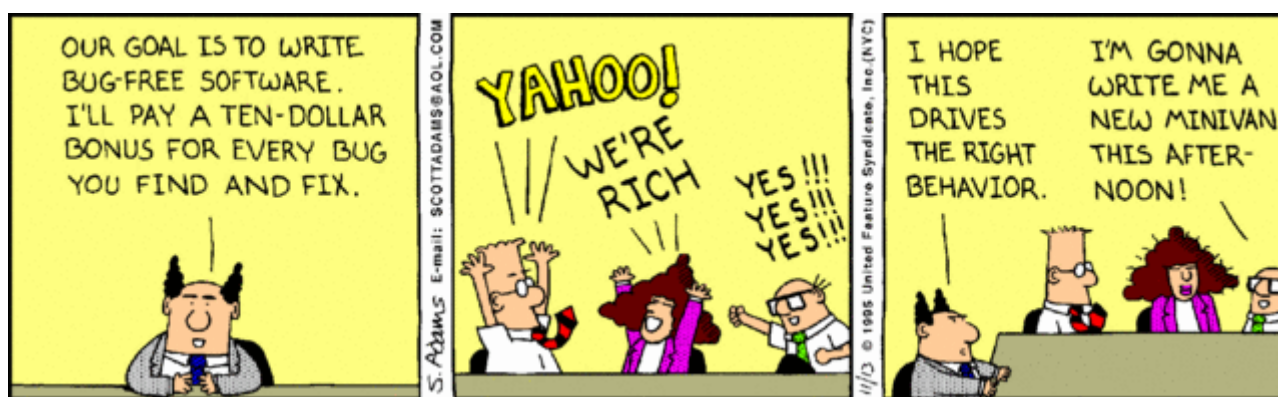
又比如，软件产品的实现方式有很大的不确定性。一个相同的功能，可以采用多种语言、框架、平台，使用各种不同的研发流程生产出来。在这种情况下，我们很难通过度量来衡量这些不同研发方法和中间过程的优劣。

**面对这样的一个复杂系统，我们不可能覆盖其全部参数。而如果这时，研发人员的利益和这个度量结果相关，那么他就很可能会通过“做数字”来欺骗度量系统。**

关于这个主题，美国著名学者罗伯特·奥斯汀（Robert Austin）写过一本书，叫作《衡量和管理组织绩效》（Measuring and Managing Performance in Organizations）。他在这本书中给出的结论是，如果你不能度量一个事物的所有方面，那就不要去度量它。否则，你将得到“做数字”的欺骗行为。

我在文稿里放了一组有名的 Dilbert 漫画。这组漫画，讲的是一个公司宣布使用 Bug 修复数量做度量，每修复一个 Bug 奖励 10 美元，消息一出，开发人员欢呼雀跃。一个程序员当场表示，当天下午就要给自己写出一辆汽车，因为他很容易就可以写出很多简单的 Bug，然后马上去修复它们。

通过这个例子，我想要和你说明的重点是：度量与绩效挂钩，结果是指标上去了，却没给软件产品带来任何好处。



备注：图片来自<https://dilbert.com/strip/1995-11-13>

而我刚刚和你举的大型公司全公司范围内推行一套度量指标的案例，正是犯了度量与绩效挂钩的错误，这种情况下，很容易出现“做数字”的不良行为。这也是使用度量时最常见的错



误，我们一定要留意。

**研发效能难以度量的第二个原因**，和上面提到的根本原因相关，但有其特殊性。很多公司有竖井（silo）存在，所以常常会把注意力放到某一两个竖井上，进行局部优化。但是，局部优化并不代表全局优化，甚至会让全局恶化。

上面那个中型公司推行质量方面指标的案例，就是在进行局部优化，不但没能提高产品质量，反而导致员工积极性受损，同时影响了团队之间的关系。

这样的问题，在按职能划分团队的公司很容易出现。因为在这样的组织划分下，更容易出现竖井，自然就更容易按照竖井来考虑小团队的表现。如果你的公司基于职能划分，一定要多加留意。

**研发效能难以度量的第三个原因在于**，度量指标一般用来度量软件产品的生产过程和产品质量，但是公司真正需要关注的是产品能否解决用户问题，也就是说能否产生用户价值。技术产品输出和用户价值输出之间的沟壑难以打通。

上面案例中那家创业公司聚焦度量开发测试指标，就是犯了这样的错误。产品质量再好，没有用户价值也是白费工夫。

当然了，研发效能难以度量还有一些其他原因，比如：

度量数据的收集难易程度不同，人们倾向拿容易收集的数据去关联效率，但事实上难以收集的数据对度量可能才真正有用。比如说代码行数容易衡量，但是用处很小。相比之下，产品用户价值的度量要困难得多，但用处却大得多。

软件开发和实践有一个滞后效应。比如，在团队中引入代码审查，在刚开始实行的时候，总体效率会出现短时间的下降，一两个月后才会逐步显现正面效果。那么，你现在要怎么度量它才好呢？

## 总结

在今天的分享中，我给出了效能度量的定义及作用，列举了三个典型的失败案例，并通过这几个例子和你详细分析了度量困难的几个主要原因。

研发效能的度量一直以来都是个难题，很多业界大佬也都发表过“研发效率不可度量”的观点。比如，马丁·福勒（Martin Fowler）在一篇叫作“[无法衡量生产效率](#)”（Cannot Measure Productivity）的文章中指出：这是我认为我们必须承认无知的地方。

周思博（Joel Spolsky）在一篇名为“[飙高音](#)”（做到最好，Hitting the High Notes）的文章中写到，衡量程序员的工作效率相当困难，原因如下：

几乎任何你能想到的指标（比如，调测过的代码行数、功能点、命令行参数的个数）都很容易被“做数字”；

我们极少要求两个程序员做完全相同的事情，所以很难获取大型项目的有价值度量数据作为参考。

那是不是说在研发软件时，我们就不能使用度量效能的方法来指导工作了呢？如果是的话，这对于软件团队管理者而言，将会是一个难以接受的事实。

这个问题的答案，我们就留到下一篇文章再揭晓吧，敬请期待。

## 思考题

你在工作中有没有经历过研发效能度量的失败案例？如果有的话，你觉得失败的原因是什么？关于度量谜题怎么解决，你有没有什么建议？

感谢你的收听，欢迎你在评论区给我留言分享你的观点，也欢迎你把这篇文章分享给更多的朋友一起阅读。我们下期再见！

---

# 研发效率破局之道

Facebook 研发效率工作法

葛俊

前 Facebook 内部工具团队 Tech Lead



新版升级：点击「 请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 01 | 效能模型：如何系统地理解研发效能？

## 精选留言 (6)

写留言



西西弗与卡夫卡

2019-08-26

写的很好，有点等不急，好想催更

展开 ∨



3



2019-08-26

问题复杂之后方法需要使用完全不一样的工具了，比如今天的深度学习使用了完全不同于传统算法的办法，虽然说可解释差，但是还是有一些好和坏的准则的。就像现在提倡okr而不是kpi，鼓励共同成长，相互尊重，分享收益等都是很积极的应对措施，还有更用心的过程控制不可缺少

展开 ∨



1





**许童童**

2019-08-26

效能到底高不高，确实很难度量，当效能和绩效挂钩时，那效能的度量标准就不能随便定了，否则很容易伤到团队。能否这样，开发一个标准功能，我们把这个功能的工作量定为1，然后在评估其它功能的时候，就以这个标准定好工作量为几，比如5，这个由大家投票，取平均值。以这种方式来度量，老师觉得怎么样。

展开 ∨



**舒偌一**

2019-08-26

公司采用任务驱动开发，度量的目标是考核程序员的效能，后来发现对互相协作带来的很大影响，后来改为针对团队，再后来就取消了。失败的原因是从影响个人或团队的收益出发而不是从提高效能出发。现在尝试要求团队在冲刺完成后提交改善意见和上次改善意见实施情况，慢慢试验中。

展开 ∨



**李双喆**

2019-08-26

问题大于解决方案哈，

展开 ∨



**Y024**

2019-08-26

如果我们理解了“因因因因因因因因果”，其实很多事情也就理解了。

