

Measuring the User Experience

Collecting, Analyzing, and Presenting Usability Metrics (Second Edition)

用户体验度量

收集、分析与呈现 (第2版)



[美] Tom Tullis Bill Albert 著

周荣刚 秦宪刚 译



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

Measuring the User Experience

Collecting, Analyzing, and Presenting Usability Metrics (Second Edition)

用户体验度量

收集、分析与呈现 (第2版)

[美] Tom Tullis Bill Albert 著

周荣刚 秦宪刚 译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

如何量化用户体验对有效提高产品的使用质量至关重要。本书详尽地介绍了如何有效且可靠地收集、分析和呈现典型的用户体验度量数据：操作绩效（正确率等）、用户体验问题（频率和严重程度）、自我报告式的满意度及生理/行为数据（眼动追踪等）。同时对“综合性量化度量数据”等问题进行了专门介绍，而且结合案例等形式对当前与用户体验相关的新内容（如用户体验对 NPS 的影响）进行了说明。

本书内容翔实，是一本值得用户体验从业人员研读的指导性书籍，同时也可以作为相关课程的参考教材。

Measuring the User Experience:Collecting, Analyzing, and Presenting Usability Metrics, Second Edition

Tom Tullis, Bill Albert

ISBN: 978-0124157811

Copyright © 2013 by Elsevier Inc. All rights reserved.

Authorized Simplified Chinese translation edition published by the Proprietor.

Copyright © 2016 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Published in China by PHEI under special arrangement with Elsevier (Singapore)Pte Ltd.. This edition is authorized for sale in the mainland of China only, excluding Hong Kong, Macau SAR and Taiwan. Unauthorized export of this edition isa violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd.授予电子工业出版社在中国大陆地区（不包括香港、澳门特别行政区以及台湾地区）出版与发行。未经许可之出口，视为违反著作权法，将受法律之制裁。本书封底贴有 Elsevier 防伪标签，无标签者不得销售。

版权贸易合同登记号 图字：01-2015-3994

图书在版编目（CIP）数据

用户体验度量：收集、分析与呈现：第2版 /（美）图丽斯（Tullis,T.），（美）艾伯特（Albert,B.）著；周荣刚，秦究刚译. —北京：电子工业出版社，2016.2

书名原文：Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics（Second Edition）

ISBN 978-7-121-27712-2

I. ①用… II. ①图… ②艾… ③周… ④秦… III.①软件设计 IV. ①TP311.5

中国版本图书馆 CIP 数据核字（2015）第 285716 号

策划编辑：孙学瑛

责任编辑：李利健

印 刷：中国电影出版社印刷厂

装 订：三河市华成印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×980 1/16 印张：21.25 字数：417 千字

版 次：2016 年 2 月第 1 版（原著第 2 版）

印 次：2016 年 2 月第 1 次印刷

定 价：99.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlt@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

献给

Tom:

我的妻子 Susan，以及我的女儿 Cheryl 和 Virginia

Bill:

我已故的父亲 Lee Albert 和已故岳母 Gita Mitra

译者序

如何提升产品的用户体验 (User Experience) 逐渐受到各相关方的重视。在“以用户为中心的设计”整个过程中, 都会涉及用户体验的评估和度量。用户体验本质上是一种主观感知, 具有明显的不确定性和模糊性, 随着产品生命周期的演变, 用户在产品认知和使用方面的需求也会发生相应的变化。用户体验自身所固有的这些属性注定了对其进行评估离不开经验型的评估方法。这种方法通常的做法就是邀请一定数量的真实用户或潜在用户, 请他们根据对产品的认知和使用经验对产品的用户体验质量进行反馈。这些反馈由用户体验从业人员根据具体的目的整理成绩效数据 (如任务完成率、求职次数和任务完成时间)、自我报告数据 (如满意度) 和用户体验问题数据 (如出现频率和优先级别), 有时候还会包括认知神经和生理方面的数据 (如眼动追踪和情绪反应)。这样, 通过对这些数据进行分析, 就可以整理出产品设计中存在的问题及后续改进的方向。不过, 也正是因为用户体验评估的方法多是经验型的方法, 不同的人在数据收集、分析和解释上会存在差异, 起到的效果也会有所不同。针对这样的问题, 本书作者系统地进行了梳理, 他们就用户体验度量方法中的方方面面进行了说明和解释, 这可以有效地帮助从业人员更加规范地对用户体验进行评估。本书的两位作者都受过严格的人因工程和心理学训练, 在用户体验方面也积累了丰富的实践经验, 本书第2版依然体现了三个特色: 第一, 围绕用户体验评估, 对几种典型的度量数据的收集和分析进行了全面整理; 第二, 操作性强, 这是一本面向用户体验从业人员的书, 作者提供了大量可读性很强的说明性样例和案例研究; 第三, 为度量数据的选择、分析和呈现提供了适当的极具针对性的理论基础, 保证了用户体验评估的科学性。

近年来, 作为一个实践导向的专业, 用户体验越来越受到业界和学校教育的重视, 其自身也在发展和更新, 比如: (1) 随着精益创业/创新的推进, 用户体验要与客户发展和客户全流程体验结合起来, 实践过程中要考虑如何基于用户的角度呈现产品的价值, 甚至还要与运营和营销结合起来考虑用户体验在提升产品成功转化

中的作用；(2) 随着移动互联网的发展，“大数据”为我们收集、分析和评估用户行为也提供了重要的途径，在用户体验实践中需要考虑“大数据”和“小数据”（如小样本的评测数据）的综合使用；(3) 就用户体验所基于的心理学基础来看，以往我们往往多以认知心理学的观点去考虑解释、评估和设计用户体验。目前心理学中“具身认知”的思想越来越受到关注，就用户体验来说，这一理论取向强调人对产品的主观体验感知会受到大脑、身体和环境的综合影响，比如：用户点头或摇头这样的身体动作和所处环境的物理特征（如温度、空间、软硬等）都会影响用户对产品体验感知的评价。这些都可以是我们在用户体验评估中需要关注的话题，本书虽然没有专门就此进行讨论，但所涉及的基本内容依然适用。

与第1版一样，现在看到的第2版依然由多位译者共同完成，初稿由周荣刚、秦宪刚、吴峰、李阳阳和薛立成完成，之后由周荣刚和秦宪刚进行校正，最后由周荣刚完成统稿和最终审校。在此感谢参与第1版翻译的其他几位译者，以及对第1版译稿予以指正的读者。本书作者在第2版更新的内容较多，翻译中的不妥之处，恳请读者批评、指正。

周荣刚

2015年11月18日

第2版前言

欢迎翻阅《用户体验度量：收集、分析与呈现》第2版。用户体验（User Experience）（经常被简称为UX）自本书第1版出版以来已发生了相当多的变化。早在2008年的时候，iPad还不存在，基于安卓系统的智能手机也尚未出现；而iPhone则刚出现不久，Pinterest和其他诸多社交网站甚至尚未被关注；谷歌Chrome浏览器也仅仅只是个传闻。我们提及这些硬件和业务，是因为它们形成了人们对类似日常技术的使用体验和期望。用户期待随手拿起一个新的应用或硬件就可以正确地使用起来。确保这些技术可以到达这样的目标正是本书的出发点。

UX涵盖了用户与产品、设备或系统交互时所涉及的所有内容。很多人似乎都把用户体验认为是一些不可被测量或量化的模糊特性。但我们认为UX是可以被度量或量化的。用来测量用户体验的工具可对下列内容进行度量（metrics）：

- 用户能否使用智能手机成功地找到他们健康计划中距离最近的医生？
- 在旅行网站上预定一个航班需要用多长时间？
- 用户在尝试登录一个新系统时犯了多少错误？
- 有多少用户可以成功地使用他们平板电脑上新安装的应用控制数字录像机去录制所有他们喜爱的电视节目？
- 有多少用户在进入一个“直奔终点”的直梯时没有首先选择自己要去的楼层，然后才发现直梯中根本就没有选择面板。
- 有多少用户在注册新业务时由于看不清被新手机电池盖着的很小的序列号而倍感受挫？
- 有多少用户在没有文字说明的情况下能够很轻易地把他们的新书架组装起来，并因此感觉愉悦？

这些都是可以通过测量获得的行为和态度方面的示例。其中，有的可能比其他的要容易测量，但它们都是可以测量的。任务成功率、任务时间、鼠标点击或敲击键盘的次数、挫折或愉悦感的自我报告式评分，甚至注意网页上某个链接的注视点

个数都是 UX 度量的例子。这些度量可以使你对用户体验形成非常有价值的见解。

为什么你需要测量用户体验呢？答案很简单：这有助于你提升用户体验。对今天的多数消费类产品、应用和网站来说，如果你不提升用户的体验，就会落后。UX 度量可以帮助你确定你相对于竞争对手所处的位置，也可以帮助你准确定位，以集中力量对需要提高的地方进行改进（即用户感觉最困惑、低效或受挫的地方）。

这是一本操作性的指导图书，而不是一本理论专著。我们主要就实际中的应用提供一些建议，比如：在什么情境下收集哪些度量、如何收集这些度量、如何使用不同的分析方法对数据进行梳理，以及如何以一种最清晰、最吸引人的方式呈现结果。我们也将与你分享实践中的一些教训，这些教训源于我们在该领域内 40 多年的经验总结。

这本书适合对提高任何类型产品用户体验有兴趣的任何人，无论这些产品是消费类产品、计算机系统、应用程序、网站，还是其他任何类型的物品。只要是人使用的产品，那么你就可以测量与使用该产品有关的用户体验。那些关注如何提高用户体验的人和那些可以从本书获益的人可以来自许多不同的专业和视角，包括可用性和用户体验（UX）专业人员、交互设计师、信息架构师、产品设计师、Web 设计师和开发者、软件开发人员、图形设计师、销售和市场营销从业人员，以及项目和产品管理人员。

与第 1 版相比，本书（第 2 版）有哪些更新呢？下面是一些最突出的内容：

- 测量情绪体验的新技术，包括手环和面部表情的自动分析。
- 眼动追踪技术的进展，包括远程基于摄像头的眼动追踪。
- 新的案例，包括 UX 领域目前正在做的事情（第 10 章案例研究全部更新）。
- 收集和分析 UX 数据的新方法和工具，包括多种在线工具。
- 通贯全书的诸多新示例。

我们希望你可以发现这本书在探求如何提高你产品的用户体验等方面对你有所帮助。我们很想听到你的成功（和失败！）。我们真的很重视许多读者针对第 1 版所给予的反馈和建议。大部分反馈都有助于促成我们在这一版本中所进行的内容上的调整和新增。你可以通过我们的网站与我们取得联系：www.MeasuringUserExperience.com。在那里，你也可以找到一些补充性材料，比如：本书多数例子中所提到的真实电子数据表格和图形，以及可以帮助你测量用户体验的相关工具等方面的信息。

致谢

首先，我们要感谢 Elsevier 的 Meg Dunkerley，我们感谢你所付出的卓越工作和适时的敦促。我们也要感谢 Joe Dumas、Bob Virzi 和 Karen Hitchcock 对文稿的审阅，你们的建议使本书的内容更加聚焦。我们也要感谢贡献案例研究的所有作者：Erin Bradner、Mary Theofanos、Yee-Yin Choong、Brian Stanton、Tanya Payne、Grant Baldwin、Tony Haverda、Viki Stirling、Caroline Jarrett、Amanda Davis、Elizabeth Rosenzweig 和 Fiona Tranquadaorgan。因为你们乐意与我们的读者分享你们的经验，使这本书更有价值。另外，我们也要感谢 Daniel Bender、Sven Krause 和 Ben van Dongen 分享你们所在机构的相关信息及目前正在使用的技术和工具。最后，我们要非常感谢富达投资和本特利大学（Bentley University）设计和可用性中心的同事，我们从你们那里学习了太多，与这样一个奇妙的 UX 研究团队共事使我们倍感荣幸。

Tom :

我要感谢我的妻子 Susan，你所给予我的支持和帮助不胜枚举，是你帮助我成为一名更好的写作者。我要感谢我的女儿 Cheryl 和 Virginia，谢谢你们一直以来的鼓励，谢谢你们对我的乏味笑话开怀大笑。

Bill :

我要感谢我的家人。Devika，你对写作的喜爱鼓舞了我；Arjun，你对数字的迷恋使我以新的方式考虑数据和度量指标。感谢我的妻子 Monika，谢谢你在我撰写本书期间所给予的支持和鼓励，没有你，我完成不了这项工作。

关于作者

Thomas S. (Tom) Tullis 是富达投资公司 (Fidelity Investments) 用户体验研究部门 (User Experience Research) 的高级副总裁, 同时也是本特利大学信息设计学院人因工程方向的兼职教授。他 1993 年加入富达用户体验, 对该公司用户体验部门的发展起了重要作用, 该部门的设备包括一个技术发展水平 (state-of-the-art) 可用性研究实验室。在加入富达投资公司之前, Tom 曾在佳能信息系统 (Canon Information Systems)、麦道 (McDonnell Douglas)、优利系统公司 (Unisys Corporation) 和贝尔实验室 (Bell Laboratories) 任职。他和富达的可用性研究团队曾被多家出版物专题介绍过, 包括《新闻周刊》(Newsweek)、*Business 2.0*、*Money*、《波士顿环球报》(The Boston Globe)、《华尔街日报》(The Wall Street Journal) 和《纽约时报》(The New York Times)。Tullis 在莱斯大学获得学士学位、在新墨西哥州立大学获得实验心理学硕士学位以及在莱斯大学获得工程心理学博士学位。他有 35 年以上的人机界面研究方面的经验, 在诸多技术期刊上发表了 50 多篇文章, 他曾在国内和国际会议上做特邀报告。同时, Tom 拥有 8 项美国专利。合作完成 (与 Bill Albert 和 Donna Tedesco 合著) 的 *Beyond the Usability Lab: Conducting Large-Scale Online User Experience Studies* 在 2010 年由 Elsevier/Morgan Kaufman 出版。Tullis 是 2011 年用户体验行业协会 (User Experience Professional Association, UXPA) 终身成就奖的获得者, 2013 年被 SIGCHI (ACM 人机交互特别兴趣组) 遴选为人机交互学会会士 (CHI Academy)。可通过 @TomTullis 关注 Tom。

William (Bill) Albert 目前是本特利大学设计和可用性研究中心的执行总监 (Executive Director), 也是本特利大学信息设计学院人因工程方向的兼职教授。在加入本特利大学之前, 他是富达投资公司用户体验部的总监, Lycos 公司的高级用户界面研究员, 也曾是 Nissan Cambridge Basic Research

的博士后研究人员。Albert 曾在 30 多个国内和国际会议上发表和报告过他的研究。2010 年，合作完成（与 Tom Tullis 和 Donna Tedesco 合著）的 *Beyond the Usability Lab: Conducting Large-Scale Online User Experience Studies*，并由 Elsevier/Morgan Kauffman 出版。他是 *Journal of Usability Studies* 的共同主编（co-Editor in Chief）。因为他在人因学和空间认知（spatial cognition）领域内的研究，Albert 获得了加州大学圣塔芭芭拉分校和日本政府所授予的奖项。他从华盛顿大学获得学士和硕士学位（地理信息系统），在波士顿大学（地理—空间认知）获得博士学位。他在 Nissan Cambridge Basic Research 完成了博士后研究。可通过 @UXMetrics 关注 Bill。

目录

第1章 引言 / 1

- 1.1 什么是用户体验 / 4
- 1.2 什么是用户体验度量 / 6
- 1.3 用户体验度量的价值 / 8
- 1.4 适用于每个人的度量方法 / 9
- 1.5 用户体验度量的新技术 / 10
- 1.6 十个关于用户体验度量的常见误解 / 11
 - 误解 1：度量需要花太多的时间而难以收集 / 11
 - 误解 2：用户体验度量要花费太多的钱 / 12
 - 误解 3：当集中在细小的改进上时，用户体验度量是没有用的 / 12
 - 误解 4：用户体验度量对我们理解原因没有帮助 / 12
 - 误解 5：用户体验数据的噪声太多 / 13
 - 误解 6：你只能相信你的直觉 / 13
 - 误解 7：度量不适用于新产品 / 13
 - 误解 8：没有度量适用于我们正在处理的问题 / 14
 - 误解 9：度量不被管理层所理解或赞赏 / 14
 - 误解 10：用小样本很难收集到可靠的数据 / 14

第2章 背景知识 / 16

- 2.1 自变量和因变量 / 16
- 2.2 数据类型 / 17
 - 2.2.1 称名数据 / 17
 - 2.2.2 顺序数据 / 18

- 2.2.3 等距数据 / 18
- 2.2.4 比率数据 / 19
- 2.3 描述性统计 / 20
 - 2.3.1 集中趋势的测量 / 20
 - 2.3.2 变异性的测量 / 22
 - 2.3.3 置信区间 / 23
 - 2.3.4 通过误差线来呈现置信区间 / 25
- 2.4 比较平均数 / 27
 - 2.4.1 独立样本 / 27
 - 2.4.2 配对样本 / 29
 - 2.4.3 比较两个以上的样本 / 30
- 2.5 变量之间的关系 / 32
 - 2.5.1 相关 / 32
- 2.6 非参数检验 / 33
 - 2.6.1 卡方检验 / 33
- 2.7 用图形化的方式呈现数据 / 35
 - 2.7.1 柱形图或条形图 / 36
 - 2.7.2 折线图 / 38
 - 2.7.3 散点图 / 40
 - 2.7.4 饼图或圆环图 / 41
 - 2.7.5 堆积条形图 / 43
- 2.8 总结 / 44

第3章 规划 / 45

- 3.1 研究目标 / 45
 - 3.1.1 形成式可用性 / 46
 - 3.1.2 总结式可用性 / 46
- 3.2 用户目标 / 47
 - 3.2.1 绩效 / 47
 - 3.2.2 满意度 / 48

3.3 选择正确的度量：10种可用性研究 / 48

3.3.1 完成一个业务 / 50

3.3.2 比较产品 / 50

3.3.3 评估同一种产品的频繁使用 / 51

3.3.4 评估导航和 / 或信息架构 / 51

3.3.5 提高知晓度 / 52

3.3.6 问题发现 / 53

3.3.7 使应急产品的可用性最大化 / 53

3.3.8 创造整体的正向用户体验 / 54

3.3.9 评估微小改动的影响 / 55

3.3.10 比较替代性的设计方案 / 55

3.4 评估方法 / 56

3.4.1 传统（引导式）的可用性测试 / 56

3.4.2 在线（非引导式）可用性测试 / 57

3.4.3 在线调查 / 60

3.5 其他研究细节 / 61

3.5.1 预算和时间表 / 61

3.5.2 参加者 / 62

3.5.3 数据收集 / 64

3.5.4 数据整理 / 64

3.6 总结 / 65

第4章 绩效度量 / 67

4.1 任务成功 / 69

4.1.1 二分式成功 / 70

4.1.2 成功等级 / 75

4.1.3 任务成功测量中存在的问题 / 78

4.2 任务时间 / 79

4.2.1 测量任务时间的重要性 / 80

4.2.2 如何收集和测量任务时间 / 80

4.2.3 分析和呈现任务时间数据 / 83

- 4.2.4 使用时间数据时需要考虑的问题 / 87
- 4.3 错误 / 89
 - 4.3.1 何时测量错误 / 89
 - 4.3.2 什么构成了错误 / 90
 - 4.3.3 收集和测量错误 / 90
 - 4.3.4 分析和呈现错误 / 91
 - 4.3.5 使用错误度量时需要考虑的问题 / 93
- 4.4 效率 / 93
 - 4.4.1 收集和测量效率 / 94
 - 4.4.2 分析和呈现效率数据 / 95
 - 4.4.2 结合任务成功和任务时间的效率 / 98
- 4.5 易学性 / 100
 - 4.5.1 收集和测量易学性数据 / 101
 - 4.5.2 分析和报告易学性数据 / 102
 - 4.5.3 测量易学性时需要考虑的问题 / 104
- 4.6 总结 / 104

第 5 章 基于问题的度量 / 106

- 5.1 什么是可用性问题 / 107
 - 5.1.1 真问题和假问题 / 108
- 5.2 如何发现可用性问题 / 108
 - 5.2.1 面对面研究 / 110
 - 5.2.2 自动化研究 / 110
- 5.3 严重性评估 / 110
 - 5.3.1 基于用户体验的严重性评估 / 111
 - 5.3.2 综合多种因素的严重性评估 / 112
 - 5.3.3 严重性等级评估系统的应用 / 113
 - 5.3.4 严重性等级评估系统的忠告 / 114
- 5.4 分析和报告“可用性问题相关的度量” / 115
 - 5.4.1 独特问题的频次 / 115

- 5.4.2 每个参加者遇到的问题数量 / 117
- 5.4.3 参加者人次 / 118
- 5.4.4 问题归类 / 119
- 5.4.5 按任务区分问题 / 119
- 5.5 可用性问题发现中的一致性 / 120
- 5.6 可用性问题发现中的偏差 / 123
- 5.7 参与者数量 / 125
 - 5.7.1 五个参与者足够 / 125
 - 5.7.2 五个参与者不够 / 127
 - 5.7.3 我们的建议 / 129
- 5.8 总结 / 129

第 6 章 自我报告度量 / 131

- 6.1 自我报告数据的重要性 / 132
- 6.2 评分量表 / 132
 - 6.2.1 Likert 量表 / 133
 - 6.2.2 语义差异量表 / 134
 - 6.2.3 什么时候收集自我报告数据 / 134
 - 6.2.4 如何收集自我报告数据 / 135
 - 6.2.5 自我报告数据收集中的偏差 / 135
 - 6.2.6 评分量表的一般指导原则 / 136
 - 6.2.7 分析评分量表数据 / 137
- 6.3 任务后评分 / 141
 - 6.3.1 易用性 / 141
 - 6.3.2 情景后问卷 (ASQ) / 141
 - 6.3.3 期望测量 / 142
 - 6.3.4 任务后自我报告度量的比较 / 143
- 6.4 测试后评分 / 147
 - 6.4.1 合并单个任务的评分 / 147
 - 6.4.2 系统可用性量表 / 148

- 6.4.3 计算机系统可用性问卷 / 150
- 6.4.4 用户界面满意度问卷 / 152
- 6.4.5 有效性、满意度和易用性的问卷 / 153
- 6.4.6 产品反应卡 / 155
- 6.4.7 测试后自我报告度量的比较 / 156
- 6.4.8 净推荐值 / 158
- 6.5 用SUS比较设计 / 159
- 6.6 在线服务 / 160
 - 6.6.1 网站分析和测量问卷 / 160
 - 6.6.2 美国客户满意度指数 / 162
 - 6.6.3 OpinionLab / 165
 - 6.6.4 在线网站调查的问题 / 167
- 6.7 其他类型的自我报告度量 / 167
 - 6.7.1 评估特定的属性 / 167
 - 6.7.2 具体元素的评估 / 170
 - 6.7.3 开放式问题 / 172
 - 6.7.4 知晓度和理解 / 173
 - 6.7.5 知晓度和有用性差距 / 174
- 6.8 总结 / 175

第7章 行为和生理度量 / 177

- 7.1 自发言语表情的观察与编码 / 177
- 7.2 眼动追踪 / 179
 - 7.2.1 如何进行眼动追踪 / 179
 - 7.2.2 眼动数据的可视化 / 181
 - 7.2.3 兴趣区 / 184
 - 7.2.4 常用眼动度量指标 / 187
 - 7.2.5 眼动分析技巧 / 189
 - 7.2.6 瞳孔反应 / 190
- 7.3 情感度量 / 191
 - 7.3.1 Affectiva 公司和 Q 传感器 / 192