



# 用户体验度量

## Measuring the User Experience

Collecting, Analyzing, and Presenting Usability Metrics



(美) Tom Tullis Bill Albert 著  
周荣刚 等译



机械工业出版社  
China Machine Press

# 用户体验度量

## Measuring the User Experience

Collecting, Analyzing, and Presenting Usability Metrics



(美) Tom Tullis Bill Albert 著  
周荣刚 等译



机械工业出版社  
China Machine Press

如何量化用户体验对有效提高产品的可用性而言至关重要。本书详尽地介绍了如何有效可靠地使用(收集、分析和呈现)几种重要的用户体验度量数据:操作绩效(正确率等)、可用性问题(频率和严重程度)、各种类型的满意度数据及生理/行为数据(如眼动跟踪)。同时对“综合性量化度量数据”和“投资回报率(ROI)”等问题进行了专门介绍。

本书内容详实,案例丰富,是一本值得该领域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

Thomas Tullis, William Albert

ISBN:978-0-12-373558-4

Copyright © 2008 by Elsevier Inc. All rights reserved.

Authorized Simplified Chinese translation edition published by the Proprietor.

ISBN:978-981-272-214-0

Copyright © 2009 by Elsevier (Singapore) Pte Ltd. All rights reserved.

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

本书简体中文版由机械工业出版社与Elsevier(Singapore)Pte Ltd.在中国大陆境内合作出版。本版仅限在中国境内(不包括中国香港特别行政区及中国台湾地区)出版及标价销售。未经许可之出口,视为违反著作权法,将受法律之制裁。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

本书版权登记号:图字:01-2009-1580

图书在版编目(CIP)数据

用户体验度量 / (美)特里斯(Tullis, T.), (美)阿伯特(Albert, B.)著;周荣刚等译. —北京:机械工业出版社, 2009.8

(UI设计丛书)

书名原文:Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics

ISBN 978-7-111-26609-9

I. 用… II. ①特… ②阿… ③周… III. 主页制作—程序设计 IV. TP393.092

中国版本图书馆CIP数据核字(2009)第040256号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:陈佳媛

北京京师印务有限公司印刷

2009年8月第1版第1次印刷

186mm × 240mm · 16.75印张

标准书号:ISBN 978-7-111-26609-9

定价:49.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

本社购书热线:(010) 68326294

# 译者序

当前，如何提升产品的可用性（Usability）逐渐受到系统设计人员越来越多的重视，其目标就是使用户在使用系统过程中获得良好的体验（User Experience）、竭力避免给用户在使用上带来困难。但是如何保证某个产品或服务能让用户获得良好的使用体验呢？通常这就需要采用“以用户为中心的设计（User-Centered Design, UCD）”，即：在产品开发之前洞悉用户及其需求并据此进行产品原型的设计，进而进行评估，并通过对评估数据的分析进行迭代式设计直至达到可用性目标。可以说，在产品整个生命周期中，都会涉及对用户体验的评估。然而，用户体验等概念本身具有强烈的模糊性，反映的多是用户认知、操作和主观感受等层面的体验。这就注定对用户体验的测量离不开经验型的评估方法。用户体验测量的方法源于实验心理学和工程心理学，通常的做法是邀请一定数量的真实或潜在用户使用产品完成一些典型的任务，由可用性工程师（Usability Engineer）收集诸如任务正确率、求助次数、任务完成时间和用户主观满意度等方面的定量数据，同时也会通过任务完成中的观察和测试之后与参加者的交流获得定性数据（如可用性问题和优点），进而通过对这些数据的统计分析获知产品设计中存在的问题、改进的方向或者总体上用户的体验情况。同时在很大程度上，也正是因为用户体验测量的方法多是经验型的方法，不同的人使用起来差异较大、起到的效果也有不同。

所以很欣喜看到作者对这一问题进行了梳理，他们就用户体验测量方法中的方方面面进行了说明和解释，这可以帮助从业人员更加规范地进行测量用户体验。本书作者都受过严格的心理学和人因学训练，具有丰富的实践经验，他们的这本书主要有这样几个特色：（1）内容详尽，他们分别对如何进行操作绩效测量、如何界定和分析测量中发现的用户体验问题、以及如何在测量中使用满意度量表进行了全面且深入的介绍，同时对诸如“如何综合多种度量指标给出总体性评价”、“卡片分类”和“投资回报率”的计算等也做了专题性整理；（2）为数据收集、分析和呈现提供了适当的理论基础，比如他们会告诉你如何使数据更为信服等；（3）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解释性样例和案例研究。

最近几年，可用性工程或用户体验专业在国内才开始受到较为普遍的关注，而用户体验测量的方法（如可用性测试，Usability Testing）被视为最有价值的可用性评估方法。把这本书介绍给国内同行，我们觉得还是很有意义的。这本书由多位译者共同完成，其中：周荣刚（北京航空航天大学）翻译了作者介绍、前言、第1章、第4章、第10章第6个案例研究和第11

#### IV

章；秦宪刚博士（中国移动研究院）翻译了第5章和第9章；徐毅斐（中国移动研究院）翻译了第3章和第10章第2个案例研究；郝学芹博士（空军第四医学研究所）翻译了第6章；陈楠（诺基亚）翻译了第10章第3和第4个案例研究；薛立成同学（中科院心理所在读硕士）翻译了第2章；李晶同学（中科院心理所在读博士）翻译了第10章第1和第5个案例；何吉波同学（美国伊利诺伊大学在读博士）翻译了第7章；第8章由薛立成和周荣刚翻译。周荣刚对全书译文进行了校正和统稿，并撰写了译者序。全部译者中有8位都受过工程心理学的研究训练。翻译本书的过程，也是我们学习的过程，任何不妥之处请读者多指正，Zhouronggang@gmail.com。

周荣刚

2009年7月于北京

## 作者简介

Thomas S. (Tom) Tullis是富达投资公司 (Fidelity Investments) User Insight的高级副总裁。他1993年加入富达，对该公司用户体验 (User Experience) 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) 可用性实验室。在加入富达公司之前，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。他和富达的可用性团队曾被多家媒体专题报道过，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 (Rice University) 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30多年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，他曾在美国和国际会议上作特邀报告。Tom还拥有8项美国专利，他也是Bentley学院信息设计中的人因学课程 (Human Factors in Information Design Program) 的教师。

William (Bill) Albert目前是富达投资用户体验部门的负责人。在加入富达之前，他是Lycos公司的高级用户界面研究员，也曾是剑桥基础研究 (Cambridge Basic Research) 的博士后研究人员。在过去几十年里，作为他研究的一部分，Bill几乎使用了所有类型的可用性度量。他发表了20多篇文章，曾在多个行业会议和学术会议上报告他的研究。因为他在人因学和空间认知 (spatial cognition) 领域内的研究，Albert获得了加利福尼亚大学 (the University of California) 和日本政府所授予的奖项。他获得了华盛顿大学的学士学位和硕士学位，获得了波士顿大学的博士学位。

# 前 言

用户体验 (User Experience, UX) 指的是用户与产品、设备或系统交互时所涉及的所有内容。很多人似乎都认为用户体验是一些不可测量和量化的特性。我们认为这是可以的。可用性度量 (usability metrics) 就是可以用来测量用户体验的工具, 比如以下这些情况:

- 用户在旅行网站上预定一个航班需要多长时间?
- 用户在登录系统时犯了多少错误?
- 有多少用户可以成功地使用他们的数字录像机去录制他们所喜爱的所有电视节目?
- 有多少用户没有先选择要去往的楼层就进了一个新的“直达终点”的电梯, 然后才发现里面没有楼层选择面板?
- 有多少用户没有注意到页面上的链接 (只要点击1次, 就可以让用户完成他们所需要的东西)?
- 有多少用户在注册新MP3播放器时由于看不清楚背面上很小的序列号而沮丧?
- 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新书架组装起来并因此感觉愉悦?

这些例子都是可以测量的行为和态度。其中一些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, 都是可用性度量的例子。这些度量可以使你给用户体验带来非常有价值的见解。

为什么需要测量用户体验? 答案是这有助于提高用户体验。对于今天的多数消费类产品或网站来说, 如果你不提高, 就会落后。可用性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 (即用户发现最为困惑、低效或受挫的领域)。

这是一本指导性质的书, 而不是一本理论专著。我们主要就实际中的应用提出一些建议, 比如: 在什么情境下收集哪种可用性度量、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又最有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多年经验的总结。

本书适合对提高产品用户体验感兴趣的任何人, 无论这些产品是消费类产品、计算机系统、应用程序、网站或者其他类型的物品。如果是供人使用的产品, 那么你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其他可以从本书获益的人, 都来

自于许多不同的专业，并具有不同的视角，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、销售和市场研究从业人员以及项目和产品管理人员。

我们希望你可以发现这本书在探求如何提高产品的用户体验等方面对你有帮助。我们很想听到你的成功和失败。你可以通过网站与我们取得联系：[www.MeasuringUserExperience.com](http://www.MeasuringUserExperience.com)。在那里，你也可以找到一些补充性材料，如本书多数例子中所提到的实际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。

\* \* \*

本书中所介绍的材料、例子和案例研究只是用来对相关信息进行说明和解释。出版商和作者不保证这些数据的合适性，也不提供基于任何特定目的的商业保证，以及不承担任何与这些材料、例子和案例研究相关的责任。

本书中所表达的观点只是作者的看法，不代表其雇主或出版商的立场和看法。



## 致 谢

首先我们要感谢Morgan Kaufmann的Diane Cerra、Denise Penrose、Mary James和Marilyn Rash。你们给了我们许多实践性的建议和适时的敦促。我们也要感谢Jim Lewis和Joyce Oghi对初稿给予了详尽而又及时的评审和建议。我们还要衷心地感谢案例研究的所有作者：Hoa Loranger、Bob Bailey、Cari Wolfson、Janice Nall、Jim Lewis、Scott Weiss、Chris Whitby、Aga Bojko和Todd Zazelenchuk。正是由于你们愿意与读者分享你们的经验，使本书增色不少、更为实用。

\*\*\*

我要感谢我的妻子Susan及女儿Cheryl和Virginia，感谢你们的支持，感谢你们容忍我在深夜和周末与计算机为伴。我还要特别感谢Susan对本书部分章节给予了审稿和修改意见。你一直在帮助我逐渐成为一个好作者。

Tom

我要感谢我的家人。Devika，你帮我写了几个句子，我希望你对写作的爱好能持续下去。Arjun，你给了一些不同度量方面的创作灵感，我知道你的好奇心会带着你走得更远。感谢我的妻子Monika，谢谢在我写这本书时你所给予我的支持、鼓励和信心，还要谢谢你在我撰写书稿的所有夜晚里耐心地辅导孩子们完成他们的家庭作业。

Bill

# 目 录

译者序	
作者简介	
前言	
致谢	
<b>第1章 引言</b> .....	1
1.1 本书架构 .....	2
1.2 什么是可用性 .....	3
1.3 可用性为什么重要 .....	4
1.4 什么是可用性度量 .....	6
1.5 可用性度量的价值 .....	8
1.6 十个关于可用性度量的常见误解 .....	9
<b>第2章 背景知识</b> .....	13
2.1 设计可用性研究 .....	13
2.1.1 选择参加者 .....	13
2.1.2 样本大小 .....	14
2.1.3 组内或组间研究 .....	15
2.1.4 平衡 .....	16
2.1.5 自变量和因变量 .....	17
2.2 数据类型 .....	17
2.2.1 称名数据 .....	18
2.2.2 顺序数据 .....	18
2.2.3 等距数据 .....	19
2.2.4 比率数据 .....	19
2.3 度量和数据 .....	20
2.4 描述统计 .....	20
2.4.1 集中趋势的测量 .....	21
2.4.2 变异性的测量 .....	22
2.4.3 置信区间 .....	23
2.5 比较平均数 .....	23
2.5.1 独立样本 .....	24
2.5.2 配对样本 .....	25
2.5.3 比较两个以上的样本 .....	25
2.6 变量之间的关系 .....	26
2.7 非参数检验 .....	28
2.8 图形化呈现数据 .....	29
2.8.1 柱状图或条形图 .....	30
2.8.2 折线图 .....	32
2.8.3 散点图 .....	34
2.8.4 饼图 .....	35
2.8.5 堆积条形图 .....	36
2.9 小结 .....	37
<b>第3章 规划可用性研究</b> .....	38
3.1 研究目标 .....	38
3.1.1 形成式可用性 .....	38
3.1.2 总结式可用性 .....	39
3.2 用户目标 .....	39
3.2.1 绩效 .....	40
3.2.2 满意度 .....	40
3.3 选择正确的度量: 十种可用性研究 .....	40
3.3.1 完成一次业务 .....	41
3.3.2 比较产品 .....	42
3.3.3 评估同一种产品的频繁使用 .....	42

3.3.4	评估导航和(或)信息架构	42	4.4	效率	73
3.3.5	提高知晓度	43	4.4.1	收集和测量效率	74
3.3.6	问题发现	44	4.4.2	分析和呈现效率数据	74
3.3.7	使要害产品的可用性最大化	44	4.4.3	结合任务成功和任务时间的效率	76
3.3.8	创造正面的用户体验	45	4.5	易学性	78
3.3.9	评估微小改动的影响	45	4.5.1	收集和测量易学性数据	78
3.3.10	比较替代性的设计	46	4.5.2	分析和报告易学性数据	79
3.4	其他研究细节	46	4.5.3	测量易学性时需要考虑的问题	80
3.4.1	预算和时间表	46	4.6	小结	81
3.4.2	评估方法	47	第5章	基于问题的度量	82
3.4.3	参加者	49	5.1	发现可用性问题	82
3.4.4	数据收集	50	5.2	什么是可用性问题	83
3.4.5	数据整理	50	5.3	如何发现可用性问题	84
3.5	小结	51	5.3.1	面对面研究	85
第4章	绩效度量	53	5.3.2	自动式研究	86
4.1	任务成功	54	5.3.3	问题的出现和结束	86
4.1.1	收集任一类型的成功度量	55	5.3.4	精细程度	86
4.1.2	二分式成功	55	5.3.5	需要多个观察者	87
4.1.3	成功等级	59	5.4	严重性等级评估	87
4.1.4	测量任务成功中的问题	61	5.4.1	基于用户体验的严重性等级评估	88
4.2	任务时间	62	5.4.2	综合多种因素的严重性等级评估	88
4.2.1	测量任务时间的重要性	62	5.4.3	严重性等级评估系统的应用	89
4.2.2	如何收集和测量任务时间	63	5.4.4	有关严重性等级评估的忠告	90
4.2.3	分析和呈现任务时间数据	65	5.5	分析和报告可用性问题	
4.2.4	使用时间数据时需要考虑的问题	67		相关的度量	90
4.3	错误	68	5.5.1	独特问题的频次	91
4.3.1	何时测量错误	68	5.5.2	每个参加者遇到的问题频次	92
4.3.2	什么构成了错误	69	5.5.3	参加者的频次	93
4.3.3	收集和测量错误	70	5.5.4	问题归类	94
4.3.4	分析和呈现错误	71	5.5.5	按任务区分问题	94
4.3.5	使用错误度量时需要考虑的问题	73	5.5.6	报告正面的问题	95

5.6	发现可用性问题时的一致性	95	6.4.7	测试后自我报告度量的比较	119
5.7	发现可用性问题时的偏差	96	6.5	用SUS比较各种设计	121
5.8	参加者数量	98	6.5.1	比较“老年人友好”网站	121
5.8.1	五个参加者足矣	98	6.5.2	比较Windows ME和Windows XP	121
5.8.2	五个参加者不够	99	6.5.3	纸质选票的比较	122
5.8.3	我们的建议	99	6.6	在线服务	124
5.9	小结	101	6.6.1	网站分析和测量问卷	124
<b>第6章</b>	<b>自我报告式的度量</b>	<b>102</b>	6.6.2	美国客户满意度指数	124
6.1	自我报告数据的重要性	102	6.6.3	OpinionLab	128
6.2	收集自我报告的数据	103	6.6.4	在线网站调查的问题	129
6.2.1	Likert量表	103	6.7	其他类型的自我报告度量	130
6.2.2	语义差异量表	103	6.7.1	评估具体的属性	130
6.2.3	什么时候收集自我报告数据	104	6.7.2	评估具体的元素	133
6.2.4	如何搜集自我报告数据	104	6.7.3	开放式问题	135
6.2.5	收集自我报告数据中的偏差	105	6.7.4	知晓度和理解	135
6.2.6	评分量表的一般指导原则	105	6.7.5	知晓度和有用度差距	136
6.2.7	分析自我报告数据	106	6.8	小结	136
6.3	任务后评分	107	<b>第7章</b>	<b>行为和生理度量</b>	<b>138</b>
6.3.1	易用性	107	7.1	观察和编码外显的行为	138
6.3.2	情景后问卷	107	7.1.1	言语行为	139
6.3.3	期望测量	108	7.1.2	非言语行为	141
6.3.4	可用性数量估计	109	7.2	需要仪器才能捕获的行为	141
6.3.5	任务后自我报告度量的比较	110	7.2.1	面部表情	141
6.4	测试后评分	113	7.2.2	视线跟踪	144
6.4.1	合并单个任务的评分	113	7.2.3	瞳孔反应	148
6.4.2	系统可用性量表	113	7.2.4	皮肤电反应和心率	150
6.4.3	计算机系统可用性问卷	114	7.2.5	其他测量方法	152
6.4.4	用户界面满意度问卷	115	7.3	小结	154
6.4.5	有效性、满意度和 易用性的问卷	117	<b>第8章</b>	<b>合并和比较度量</b>	<b>156</b>
6.4.6	生成反应卡	118	8.1	单一可用性分数	156
			8.1.1	根据预定目标合并度量	156

8.1.2	根据百分比合并度量	157	10.1.5	个人简介	198
8.1.3	根据z分数合并数据	161	10.2	语音识别的IVR系统的 可用性评估	198
8.1.4	使用SUM: 单一可用性度量	163	10.2.1	方法	198
8.2	可用性记分卡	165	10.2.2	结果: 基于任务的测量	199
8.3	与目标和专家绩效比较	167	10.2.3	PSSUQ	199
8.3.1	与目标比较	167	10.2.4	用户评论	200
8.3.2	与专家绩效比较	168	10.2.5	可用性问题的	200
8.4	小结	169	10.2.6	样本量的合适性	201
第9章	专题	171	10.2.7	基于用户行为和评价的建议	202
9.1	在线网站数据	171	10.2.8	讨论	203
9.1.1	服务器端的日志	171	10.2.9	作者简介	203
9.1.2	点击率	172	10.2.10	参考文献	203
9.1.3	弃用率	174	10.3	CDC.gov网站的再设计	203
9.1.4	A/B研究	175	10.3.1	可用性测试级别	204
9.2	卡片分类数据	176	10.3.2	基线测试	205
9.2.1	开放式卡片分类数据的分析	177	10.3.3	任务场景	206
9.2.2	封闭式卡片分类数据的分析	181	10.3.4	定性分析的发现	206
9.3	可达性数据	183	10.3.5	绘制线框图与首次点击测试	207
9.4	投资回报率数据	186	10.3.6	最终原型测试	209
9.5	六西格玛	188	10.3.7	结论	211
9.6	小结	190	10.3.8	个人简介	212
第10章	案例研究	192	10.3.9	参考文献	212
10.1	如何又快又省地对网站 进行再设计	192	10.4	可用性基准化: 手机 音乐和视频	212
10.1.1	阶段1: 对竞争对手的网站 进行可用性测试	192	10.4.1	项目目标和方法	213
10.1.2	阶段2: 对三种不同的设计 概念进行测试	194	10.4.2	定性和定量数据	213
10.1.3	阶段3: 对单独的设计 方案进行测试	197	10.4.3	研究范围	213
10.1.4	总结	198	10.4.4	对比分析	213
			10.4.5	调研实施: 参加者的数量	214
			10.4.6	招募参加者	214
			10.4.7	数据收集	215

10.4.8	任务完成时间	215	10.6.1	OneStart: 印地安纳大学的 门户网站项目	228
10.4.9	任务完成与失败	215	10.6.2	设计和执行研究	229
10.4.10	尝试次数	215	10.6.3	分析和解释结果	229
10.4.11	感知性度量	216	10.6.4	分享发现和建议	230
10.4.12	定性发现	216	10.6.5	对影响进行反思	231
10.4.13	定量发现	216	10.6.6	结论	232
10.4.14	总结性发现和SUM度量	217	10.6.7	致谢	232
10.4.15	数据处理及可视化	217	10.6.8	个人简介	232
10.4.16	讨论	217	10.6.9	参考文献	233
10.4.17	基线调整和后续工作	218	<b>第11章 推进</b>		234
10.4.18	个人简介	218	11.1	推销可用性和度量的作用	234
10.4.19	参考文献	219	11.2	从小事做起, 努力慢慢做起来	235
10.5	测量“药品标签设计和相似性 对药剂师操作行为的影响”	219	11.3	确保你有时间和经费	236
10.5.1	测试参加者	220	11.4	早计划, 常计划	236
10.5.2	仪器与设备	220	11.5	给你的产品确定基线	237
10.5.3	实验刺激	221	11.6	挖掘你的数据	238
10.5.4	实验程序	223	11.7	说商业性的语言	238
10.5.5	数据分析	223	11.8	呈现你的置信程度	239
10.5.6	结果和讨论	224	11.9	不要误用度量	240
10.5.7	个人简介	227	11.10	简化你的报告	240
10.5.8	参考文献	227	<b>参考文献</b>		242
10.6	重视度量	227			

# 第1章 引言

这本书的目的主要是向读者介绍可用性度量 (usability metrics) 如何作为一个有效的工具成功地对任何产品进行用户体验评估。在考虑可用性度量时, 有些人往往会被复杂的公式、似是而非的研究结果和高级的统计方法所“吓倒”。而我们期望能使大量的研究“非神秘化”, 并把重点集中在可用性度量的实践应用上。所以, 我们将带领你通过逐步分解的方法收集、分析和呈现可用性的度量; 帮助你为每一个具体的情境和实际应用选择合适的度量, 并会指引你在不超出预算的情况下, 使用这些度量获得可靠而又可控的结果; 同时我们还会告诉你一些用来分析大多数可用性度量的准则和小提示, 并提供不少实例来说明如何以简捷而有效的方式向他人呈现可用性度量的结果。

我们旨在使本书成为实践类的书, 即如何对产品的可用性进行测量。我们不会给你许多公式, 事实上, 这个领域中的公式本来就少。书中所涉及的统计也是相对有限的, 相应的计算用Excel或其他常见的软件包或Web工具就可以轻松地完成。所以我们旨在仅仅给你提供一些评估产品可用性时所需要的工具, 而不是罗列那些“望而生畏”的不必要的细节。

这本书兼备产品和技术导向, 我们所阐述的可用性度量在实践中可用于任何类型的产品和任何类型的技术, 这是可用性度量最为显著的特性之一: 它们不仅仅适用于网站或任何其他单一的技术。比如说, 无论你评估的是网站、脚踏车还是烤箱, 任务成功(率)和满意度都是同样有效的。更多的高技术产品(如网站、手机、软件、消费类电子等)要有一些特别的关注, 因为一般来说它们更为复杂, 但基本前提和其他产品都是一致的。

可用性度量的“半衰期”(half-life)比任何一个特定的设计或技术都要大得多。无论技术上发生了多大的变化, 这些度量在本质上都是一样的。随着用于测量可用性的新技术的发展, 有些度量也会跟着变化, 但被测现象的根基是没有变化的。视线跟踪(eye tracking)就是很好的例子。许多研究者希望有一个方法能获知用户注视屏幕的精确位置。如今, 随着视线跟踪技术的最新进展, 测量变得越来越简单, 也更为准确。

那么为什么我们要写这本书呢? 当然不缺少人因学(Human Factors)、统计、实验设计和可用性方法等方面的书籍, 而且其中有些甚至包括了一些更为常用的可用性度量。那么一本专门聚焦于可用性度量的书有意义吗? 显然, 我们认为是有意义的。以我们(谦卑的)观点来看, 这本书对于丰富可用性出版物来说有五方面的独特贡献:

- 我们以全面的视角来审视可用性度量。目前, 还没有其他书汇总了这么多不同的可用

性度量。对于你可能使用的几乎所有类型的可用性度量，我们在（数据）收集、分析和呈现等方面都提供了详细的介绍。

- 本书采取实用的方法。我们假设你有兴趣把可用性度量作为自己工作的一部分。在行文时，我们不会纠缠于细节而浪费你的时间。我们希望你每天都能很轻松地使用这些度量。当然如果你有兴趣了解相应的理论背景，我们也可以给你提供一些额外的信息资源。
- 在有关可用性度量方面的正确决策上，我们会提供一些帮助。可用性专业相关工作中最为困难的一个方面是要决定是否需要收集度量的数据，如果是，需要收集哪些度量数据。我们会指引你如何通过一个合适的决策过程找到合适你测量情境的正确度量。
- 我们提供了不少实例，这可以有助于我们理解可用性度量如何被应用于不同的组织，以及这些度量又如何用来对可用性问题进行诠释。我们也提供深度案例研究，以帮助你判断如何更好地使用可用性度量所揭示出来的信息。
- 我们阐述了可用于任何产品或技术的可用性度量。我们处理这个问题的角度很广，因此这些可用性度量也有助于你的职业发展，即便是在技术和产品发生了变化的情况下，也同样如此。

## 1.1 本书架构

这本书主要由三个部分组成。第一部分（第1~3章）介绍了可用性度量的背景信息。这部分的内容主要面向那些对可用性、数据分析或统计不太熟悉的读者。

- 第1章对可用性和可用性度量做了一个概述性的介绍。我们对用户体验（User experience）、可用性和多种类型的可用性度量进行了定义，对测量用户体验的价值进行了讨论；同时还消除了一些常见的有关可用性度量的“诟病”或误解。
- 第2章包括了可用性数据和一些基本统计概念方面的背景信息。我们会一步一步地向你展示如何使用不同度量确立一项可用性研究的过程，同时也介绍了相应于不同可用性方法而进行的常规统计流程。
- 第3章则集中介绍如何计划一项可用性研究，包括定义参加者目标和研究目标及为各种各样的情境选择合适的度量。

第二部分（第4~9章）对5种通用的可用性度量类型以及一些专门的不属于任何某种单一类型的专题进行了回顾。对于每一种度量，我们阐释了该度量是什么、什么时候该用和什么时候不该用等问题。我们向你介绍了如何收集以及分析和呈现这些数据的多种方法。同时我们还提供了一些例子来说明每一种度量在实际可用性研究中如何应用。

- 第4章涵盖了多种绩效度量类型，包括任务成功（task success）、任务时间（time on task）、错误（error）、效率（efficiency）和易学性（ease of learning）。这些度量由于测量的是用户行为的不同方面，因此被归纳于绩效的“伞下”。



- 第5章着眼于可用性问题的测量上。通过测量问题频次、严重程度和问题类型，可用性问題容易被量化。对诸如多大的样本量才合适及如何可靠地获取可用性问題等这样一些有争论的问题，我们也进行了讨论。
- 第6章集中介绍了自我报告度量 (self-reported metrics)，如满意度 (satisfaction)、期望 (expectation)、易用性等级评分 (ease-of-use ratings) 和知晓程度/意识 (awareness)。自我报告度量基于用户自身所分享的经验，而不是可用性专家所测查的用户真实行为。
- 第7章专注于行为和心理学度量。这些度量包括视线跟踪、面部表情和多种压力测量。所有这些度量获取的是有关个体与用户界面发生交互时机体所表达出来的行为反应。
- 第8章讨论的是如何综合不同类型的度量而获得一个新的度量。有时这有助于获得产品可用性的总体性评价。总体性的评价可以通过这样的方式实现：把不同类型的度量合成一个单一的可用性分数，或以可用性计分卡的形式总结这些度量，亦或把这些度量结果和专家绩效进行比较。
- 第9章介绍的是专题，即我们认为重要、但不能简单归为上述5种类型中任何一种的主题。这些主题诸如针对在线网站所进行的A/B测试、卡片分类 (card-sorting) 数据、六西格码 (six sigma)、可达性数据 (accessibility data) 和投资回报 (return on investment, ROI) 等。

第三部分 (第10~11章) 对可用性度量如何被应用于实践进行了介绍。在这部分中，我们突出的是：可用性度量在不同类型的组织内是如何进行实际应用的，以及如何在一个组织内提升和推广可用性度量的应用。

第10章介绍了六个案例研究。每个案例研究对不同类型的可用性度量如何使用及数据如何收集、分析和呈现结果等进行了回顾。这些案例均由可用性从业人员撰写，他们来自于多种类型的组织，包括咨询、政府、工业和非盈利机构/教育。

第11章罗列了10个可以帮助你组织内推广使用可用性度量的步骤。本章讨论了：可用性度量如何适合于不同类型的组织，有哪些实践的技巧可以使可用性度量在你的组织内起到效果，以及一些获得成功的处方。

## 1.2 什么是可用性

在我们试图测量可用性之前，我们应该弄清楚什么是可用性和什么不是可用性。可用性的定义有许多，或许这个领域的每个人都可以有自己的一个定义。这里，我们主要列出三种定义。

国际标准化组织 (ISO 9241-11) 确定了可用性的三个方面，把可用性定义为“特定使用情境下，特定的用户完成特定的目标时，产品所表现出来的效果、效率和令人满意程度”。

可用性专业协会 (Usability Professionals Association, UPA) 更多从产品开发过程的角度