

# 用户体验调查问卷手册

在你的项目中成功应用UEQ所需要知道的一切

作者。Martin Schrepp博士

第8版 (31.12.2019)

## 简介

应用用户体验问卷调查（UEQ）所需的知识目前被分割成几个独立的出版物。本手册的目标是将所有这些知识汇集到一个文件中。这将使从业者更容易在他们的评估项目中应用UEQ。

我们把重点放在最重要的事实上，以保持文件的简短（因为每增加一页都会大大减少阅读的人数）。我们引用了一些出版物，供那些想深入研究这个问题的人参考。除本手册外，请查看网站[www.ueq-online.org](http://www.ueq-online.org)，了解有关UEQ的新进展和出版物。

## UEQ的建设

UEQ的原始德国版本创建于2005年。为了确保所构建的量表的实用性，我们采用了数据分析的方法，也就是说，这些量表来自于有关更大的项目库的数据。每个量表描述了互动产品的一个独特的质量方面。

在与可用性专家的头脑风暴会议上，创建了一个由229个与用户体验有关的潜在项目组成的初始项目集。然后，通过专家评估，这个项目集被缩减为80个原始版本的调查问卷。

在一些关注互动产品质量的研究中，包括统计软件包、手机通讯录、在线合作软件或商业软件，都是用这个80个项目的原始版本来收集数据的。总共有153人回答了原始版本

的80个项目。最后，通过主成分分析，从这个数据集中提取了6个UEQ量表和代表每个量表的项目。

这些项目具有语义差异的形式，也就是说，每个项目由两个含义相反的术语表示。每个项目的术语顺序是随机的，即一个量表的一半项目从正面术语开始，另一半项目从负面术语开始。

术语。我们使用七级量表，以减少众所周知的这类项目的中心倾向偏差。

一个项目的例子是。

吸引人的    o o o o o o o    吸引人的

这些项目的标度从-3到+3。因此，-3代表最消极的答案，0代表中性答案，而+3代表最积极的答案。

UEQ量表的一致性和它们的有效性（即量表真正测量了它们想要测量的东西）在11次可用性测试中进行了调查，总共有144名参与者，在一次在线调查中，有722名参与者。这些研究的结果显示了足够高的量表一致性（通过Cronbach's Alpha测量）。此外，一些研究表明量表具有良好的构架效度。

如果你想知道关于UEQ的构建和验证的更多细节，请参阅。

*Laugwitz, B., Schrepp, M. & Held, T. (2008). 用户体验调查问卷的构建和评估。在。Holzinger, A. (Ed.): usab 2008, Lncs 5298, 63-76.*

## 规模结构

UEQ包含6个量表，共26个项目。

- **吸引力。**对产品的总体印象。用户是喜欢还是不喜欢该产品？
- **透彻性：**是否容易熟悉产品？是否容易学习如何使用该产品？
- **效率。**用户能否在不付出不必要的努力的情况下解决他们的任务？
- **可靠性。**用户是否感觉到对互动的控制？
- **刺激性。**使用该产品是否令人激动和有动力？
- **新颖性：**产品是否具有创新性和创造性？产品是否吸引了用户的兴趣？

吸引力是一个纯价值维度。**敏锐性、效率和可依赖性**是实用的质量方面（以目标为导向），而**刺激性和新颖性**是享乐的质量方面（非以目标为导向）。

吸引力量表有6个项目，其他量表都有4个项目。图1显示了UEQ的假定量表结构和每个量表的英文项目。

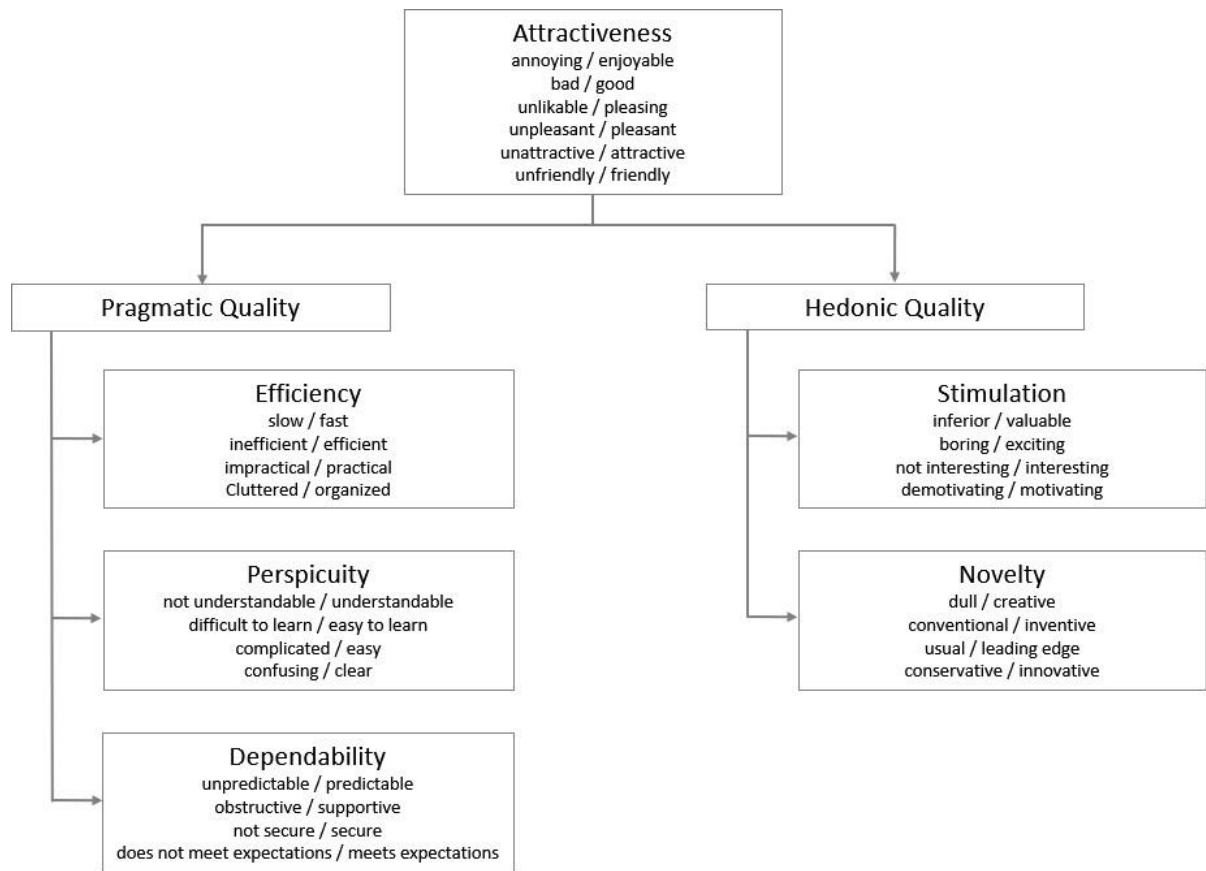


图1：UEQ的假定规模结构。

## 典型的应用场景

一些不同的研究问题可以通过UEQ对产品用户体验的定量测量来回答。

### 比较两个产品的用户体验

一个典型的情况是将一个既定的产品版本与一个重新设计的版本进行比较，以检查新版本是否具有更好的用户体验。另一种情况是将一个产品与市场上的直接竞争对手进行比较。

通过对两个UEQ测量值的统计比较，产品可以相对容易地进行比较。因此，两个产品或两个产品版本的UEQ评价是在每个UEQ量表的量表平均值的基础上进行比较。

让我们看一个例子。图2显示了两个假想的产品版本A（新）和B（旧）的比较。

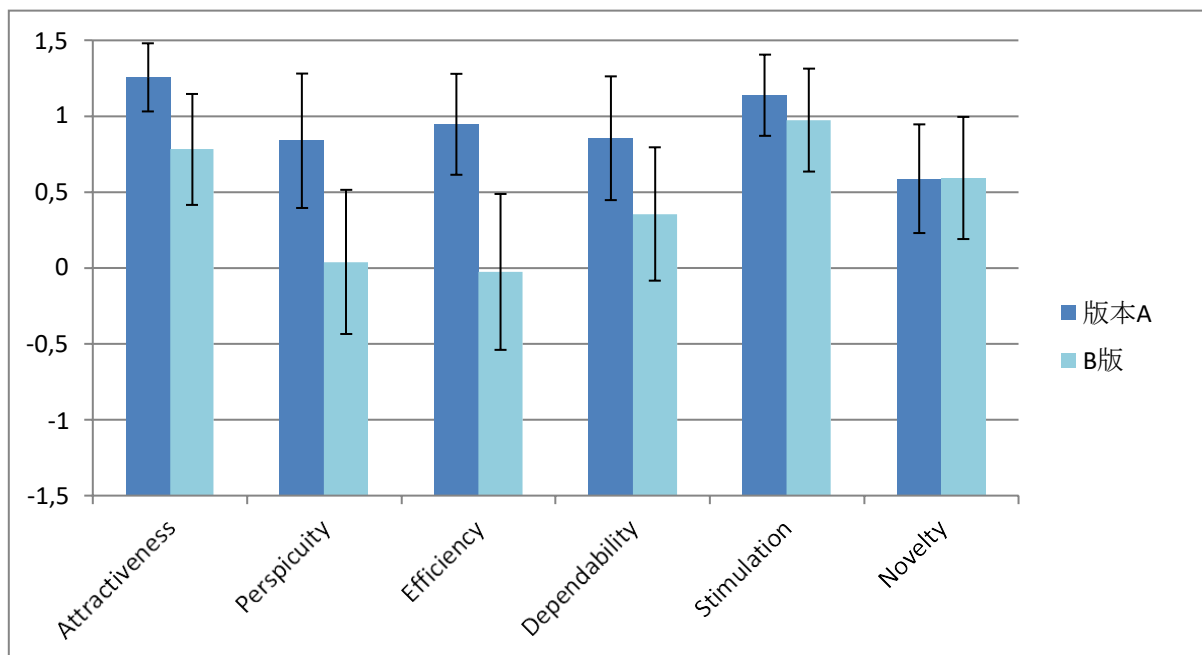


图2：两个假想的产品版本的比较。

正如我们所看到的，新版本A在所有的量表上都显示，除了新奇度的数值大致相同外，其他的数值都比旧版本B好。然而，如果你想就这个结果得出结论（特别是如果你的样本很小），你必须检查差异是否显著。

误差条代表刻度平均值的95%置信区间。它们是什么意思？假设你可以在相同的条件下无限次地重复一项评估。当然，由于一些随机的影响，你不会在每次重复中测量出完全相同的量表平均值。95%的置信区间就是这些假设的重复中95%的量表平均值所在的区间。因此，它显示了你的测量有多精确。

如果两个测量值的置信区间不重叠，那么差异在5%的水平上是显著的。在我们上面的例子中，这只适用于规模效率。但相反的结论并不正确，也就是说，如果置信区间重叠，差异仍然可以是显著的。因此，做一个显著性检验是有意义的（假设不等方差的简单双样本t检验可以用Excel *UEQ\_Compare\_Products\_Version<x>.xlsx*（<x>是版本号，随Excel的每个新版本而变化）进行，可在[www.ueq-online.org](http://www.ueq-online.org)，或者可以用每个统计包轻松完成）。

如果你将一个新版本和一个已经使用过的版本进行比较，你应该尽量在用户熟悉了新版本后再收集数据。如果你在新版本推出后直接开始收集数据，由变化本身引起的问题（即在新版本中的工作方式与旧版本不同，用户对此感到恼怒或生气，即使新版本更好）可能会严重影响你的结果。

## 测试一个产品是否有足够的用户体验

产品是否满足关于用户体验的一般期望？用户的这种期望是由他们经常使用的产品形成的。

有时，考虑到比例尺的含义，这个问题的答案立刻就很清楚了，比如下面的例子。

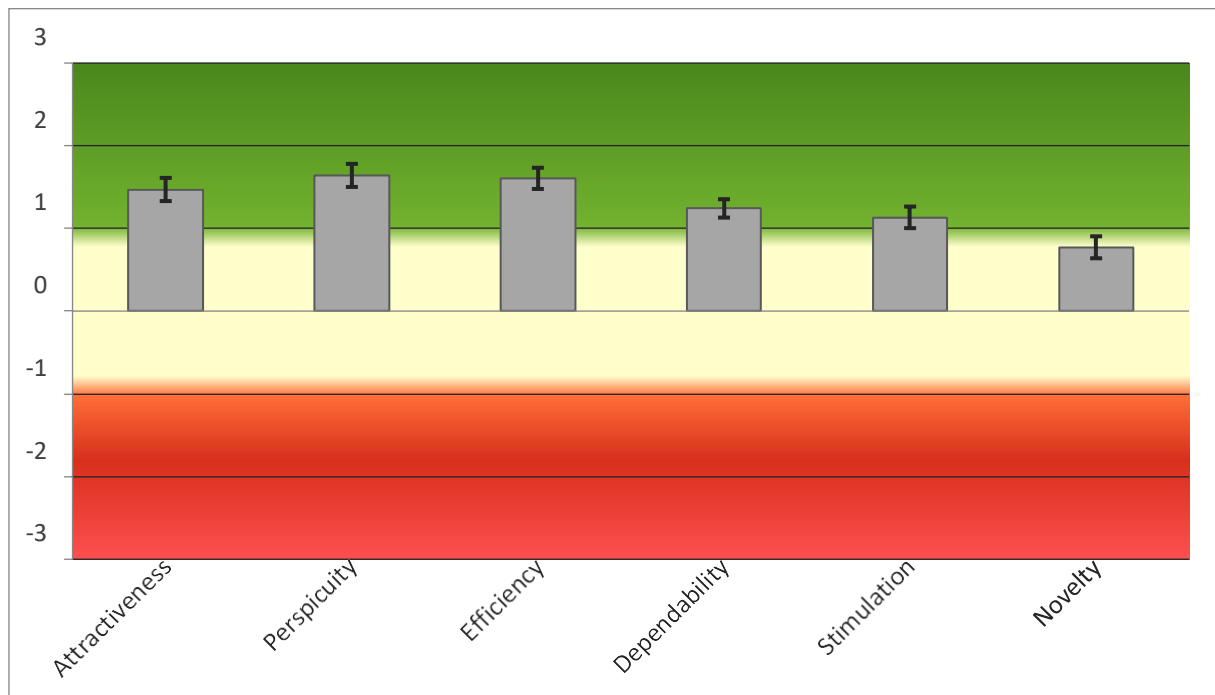


图3：具有良好效果的产品实例。

这里很明显，所有的量表都显示出了极其积极的评价。对量表含义的标准解释是，-0.8和0.8之间的数值代表对相应量表的中性评价，>0.8的数值代表积极评价，<-0.8的数值代表消极评价。

量表的范围是在-3（糟糕透顶）和+3（极好）之间。但在实际应用中，一般只观察到有限范围内的数值。由于对具有不同意见和答案倾向的不同人的平均值的计算（例如，避免极端的答案类别），极不可能观察到高于+2或低于-2的数值。

但是在典型的评估中，事情并不那么明显。为了更好地了解一个产品的质量，有必要将产品的测量用户体验与其他成熟产品的结果进行比较，例如从一个包含不同典型产品的基准数据集中进行比较。

UEQ提供了这样一个基准，其中包含了目前UEQ的452个产品评估的数据（所有评估中共有20190名参与者）。该基准每年更新一次，所以请确保你总是从网站[www.ueq-online.org](http://www.ueq-online.org)，下载最新版本的Excel-工具。

该基准将一个产品分为5个类别（每个尺度）。

- 优秀：在10%的最佳结果范围内。
- 好：在基准数据集中有10%的结果比较好，75%的结果比较差。
- 高于平均水平：基准中25%的结果优于被评估产品的结果，50%的结果较差。
- 低于平均水平。基准中50%的结果优于被评估产品的结果，25%的结果较差。
- 不好。在25%的最差结果范围内。

Excel工具中的基准图显示了你所评估的产品的用户体验质量如何。

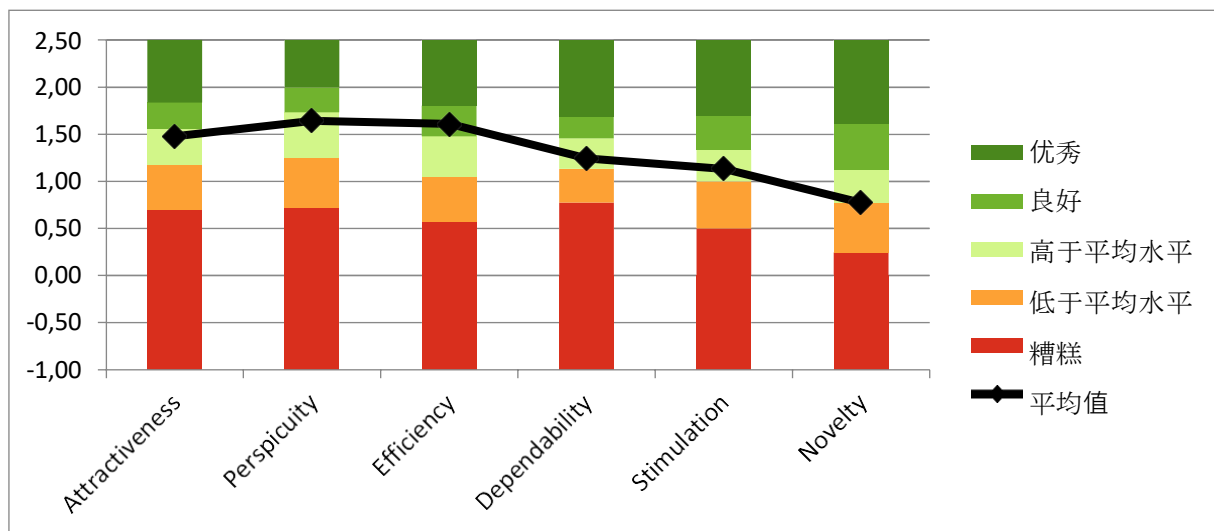


图4：一个假设的产品的基准图。

## 确定需要改进的领域

为了改善产品的用户体验，应该改变什么？这个问题不能通过对用户体验的定量测量来直接回答。为了回答这个问题，需要将产品特性与测量结果联系起来。

然而，通过像UEQ这样的问卷调查，我们至少可以对那些改进会产生最大影响的领域做出有根据的猜测。对于一个被评估的产品，UEQ显示了6个被测量的用户体验质量的模式。从这个模式中，我们至少可以做出一些假设，来寻找改进的地方。

关于UEQ的应用场景的更多细节和一些例子载于。

Schrepp, M.; Hinderks, A. & Thomaschewski, J. (2014). 在不同的评估场景中应用用户体验调查问卷 (UEQ)。In: Marcus, A. (Ed.): 设计、用户体验和可用性。设计用户体验的理论、方法和工具。计算机科学讲义》，第8517卷，383-392页，Springer国际出版社。

关于创建该基准的细节可在以下文件中找到。

*Schrepp, M.; Hinderks, A. & Thomaschewski, J. (2017). 构建用户体验问卷 (UEQ) 的基准。International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4(4), 40-44.*

## 翻译成其他语言

UEQ已经被翻译成多种语言（在本手册最后一次更新时为21种）。所有可用语言版本的完整和实际列表可在[www.ueq-online.org](http://www.ueq-online.org)。如果你的语言没有可用的版本，那就简单地创建一个吧！

以下过程几乎被用于目前所有的翻译工作中，通常有助于在翻译过程中获得良好的结果。

1. 从现有的语言版本中选择一个作为翻译的基础。我们在下文中称这个版本的语言为源语言。
2. 将说明和项目的术语翻译成你的语言（目标语言）。记住，每个项目由两个术语组成，代表一个语义维度的对立面。确保翻译后也是如此。
3. 找到其他人将目标语言的项目翻译回源语言。重要的是，译入目标语言和译回目标语言的工作要由两个独立的人完成。
4. 分析你的翻译所依据的UEQ版本与翻译回源语言后的版本之间的偏差。
5. 如果可能的话，用新的译本收集一些数据（最好是几个产品的数据），检查Cronbach alpha值是否足够高。如果这是不可能的，请将你的译文发布在[www.ueq-online.org](http://www.ueq-online.org)，其他人可以使用它并贡献他们的数据。
6. 不要忘记把你的翻译发给我们！我们将在UEQ网站上以作者身份发表，并附上您的名字（如果您同意，不写作者名字也可以发表）。即使它不完美，这也是一个好的开始，其他人可以使用它并分享他们的数据，然后可以用来改进语言版本。

关于翻译过程和检查语言版本质量的方法的更多细节，请见。

*Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S. & Thomaschewski, J. (2013)。 高效测量互动产品的用户体验--如何使用用户体验调查问卷 (UEQ)。 例子。 西班牙语语言版本。 International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 2(1), 39- 45.*

## 在线申请

UEQ很短，可以在线申请。这通常在收集数据时节省了很多精力。然而，请考虑到在网上



研究中，你可能会较高比例的人不认真填写问题。如果参与者因填写问卷而获得奖励（例如参与抽奖），情况就更是如此。

过滤掉可疑反应的一个简单策略是基于这样一个事实，即一个量表中的所有项目或多或少都是测量相同的质量方面。因此，对这些项目的回答至少不应该有太大差别。

作为一个例子，请看以下对量表"透彻性"项目的回答。

不可理解	○ ○ ○ ○ ○ x ○	可以理解
易学	○ ○ ○ ○ ○ ○ x	难学
复杂的	○ ○ ○ ○ x ○ ○	容易
清楚	○ ○ ○ ○ ○ x ○	混乱的

很明显，这些答案并不十分一致。如果将它们转为负数

(1) 因此，一个量表中所有项目的最佳答案和最差答案之间的距离很高，说明答案的行为是不一致的或随机的。

如果只在一个量表上出现如此高的距离，这并不能成为排除被调查者答案的真正理由，因为这种情况也可能是由于回答错误或对一个项目的简单误解造成的。如果这种情况发生在几个量表上，那么很可能是被试至少有一部分没有认真回答问卷。

因此，一个简单的启发式方法是，如果对于2个或3个量表（你可以决定如何严格地应用这一规则），对量表中的一个项目的最佳反应和最差反应之间的距离超过3，则认为该反应是可疑的。

这个启发式方法也在Excel工具中的一个工作表中实现。关于这个启发式方法的推导和经验评估的细节，可以在以下文章中找到（目前只有德语版本）。

Schrepp, M. (2016). Datenqualität bei Online-Fragebögen sicherstellen. S. Hess & H. Fischer (Hrsg.): Mensch und Computer 2016 - Usability Professionals. <http://dx.doi.org/10.18420/muc2016-up-0015>.

## 应用UEQ作为可用性测试的一部分

通常情况下，UEQ是作为经典的可用性测试的一部分，用来收集一些关于参与者对用户体验印象的定量数据。把问卷交给参与者的最佳时间点是在他们完成测试任务之后。如果参与者在与测试者就产品进行了长时间的讨论之后才填写问卷，这将影响测试结果。UEQ的目标是抓住用户对产品的直接印象。因此，尽量在与参与者讨论之前得到UEQ的答案。

一些参与者可能不熟悉调查问卷的特殊项目格式。当你把问卷交给参与者的时候，要提到这是一份经过科学评估的问卷，用来测量用户体验。这将提高答案的质量和一致性。

## 如何使用Excel工具？

Excel-工具的目标是使UEQ数据的分析对你来说尽可能的简单。您只需在Excel *UEQ\_Data\_Analysis\_Tool\_Version<x>.xlsx*中的相应工作表中输入数据，然后所有相关的计算（如果您想比较两种产品，显著性检验除外，这里您需要使用Excel *UEQ\_Compare\_Products\_Version<x>.xlsx*）将自动完成。

Excel工具包含解释不同计算方法的注释，因此我们不需要在此详述。请确保你总是使用[www.ueq-online.org](http://www.ueq-online.org)，因为UEQ团队试图根据用户的反馈不断改进这个工具，所以你总是使用最实际的Excel工具版本。

## 如何解释数据

### 误差条是什么意思？

比例尺的误差条显示比例尺平均值的95%置信区间。误差条与量表平均值一起显示在Excel工具的相应图表中。

假设你可以在相同的条件下经常重复一项评估。当然，由于一些随机的影响，你不会在每次重复中测出完全相同的量表平均值。误差条描述的是这些重复中95%的量表平均值所处的区间。因此，它显示了你的测量有多精确。误差条的大小取决于样本量（参与者越多，误差条通常越小）和不同参与者的同意程度（同意程度越高，即答案越相似，误差条越小）。

因此，如果置信区间比较大，你应该仔细解释你的结果。在这种情况下，你的测量可能不是很准确。通常情况下，这是因为样本量太小，也就是说，如果可能的话，你应该多收集一些数据。Excel工具中包含一个标签*Sample\_Size*，它可以帮助你估计你需要收集多少数据才能在估计尺度平均值时达到一定的精度。

### Cronbach-Alpha值是什么意思？

Excel工具中的一张工作表显示了UEQ六个量表的Cronbach-Alpha系数（ $=n * r/1 + (n - 1) * r$ ，其中 $r$ 是一个量表中项目的平均相关性， $n$ 是一个量表中项目的数量）。阿尔法系数是衡量量表一致性的一个指标，即它表明量表中的所有项目都是测量类似的结构。

没有明确的规则来描述Alpha-系数应该有多大。一些经验法则认为数值大于0.6或大于0.7是一个足够的水平。如果其中一个量表的Alpha-值太小，那么你应该仔细解释这个量表。

造成Alpha值小的原因主要有两个。首先，较多的参与者可能对量表中的某些项目产生误解。例如，"不安全/安全"这个项目与量表"可依赖性"有关。它通常被解释为：互动是可以被用户保存和控制的。在社交网络的背景下，这个项目可能被误解为"我的数据安全吗？"，也就是说，这个项目在这个背景下得到了一个非本意的含义。这

降低了与量表中其他项目的相关性，从而降低了Alpha值。第二，一个量表可能与某个产品无关。在这种情况下，人们的答案会不太一致，因为参与者在判断用户体验质量方面会有问题，而这个方面对于被调查的产品来说并不重要。这也会导致低Alpha值。从这些例子中可以看出，在这种情况下，应该谨慎地解释量表的平均值。

请注意，Alpha系数对抽样效应相当敏感。因此，如果你只有一个小的样本（例如20到40个参与者），Alpha值低可能是抽样效应的结果，不一定表明量表一致性有问题。在这种情况下，最好忽略Alpha，直接检查（见下文）量表内单个项目的平均值，以便对那些可能是关键的项目得出结论，即被研究对象误解的项目。

如果一个量表的Alpha值很小，看一下单个项目的平均值是有意义的。在那里，你有时可以直接看到某些项目是否没有以通常的方式进行解释。下图就是这样一个例子。

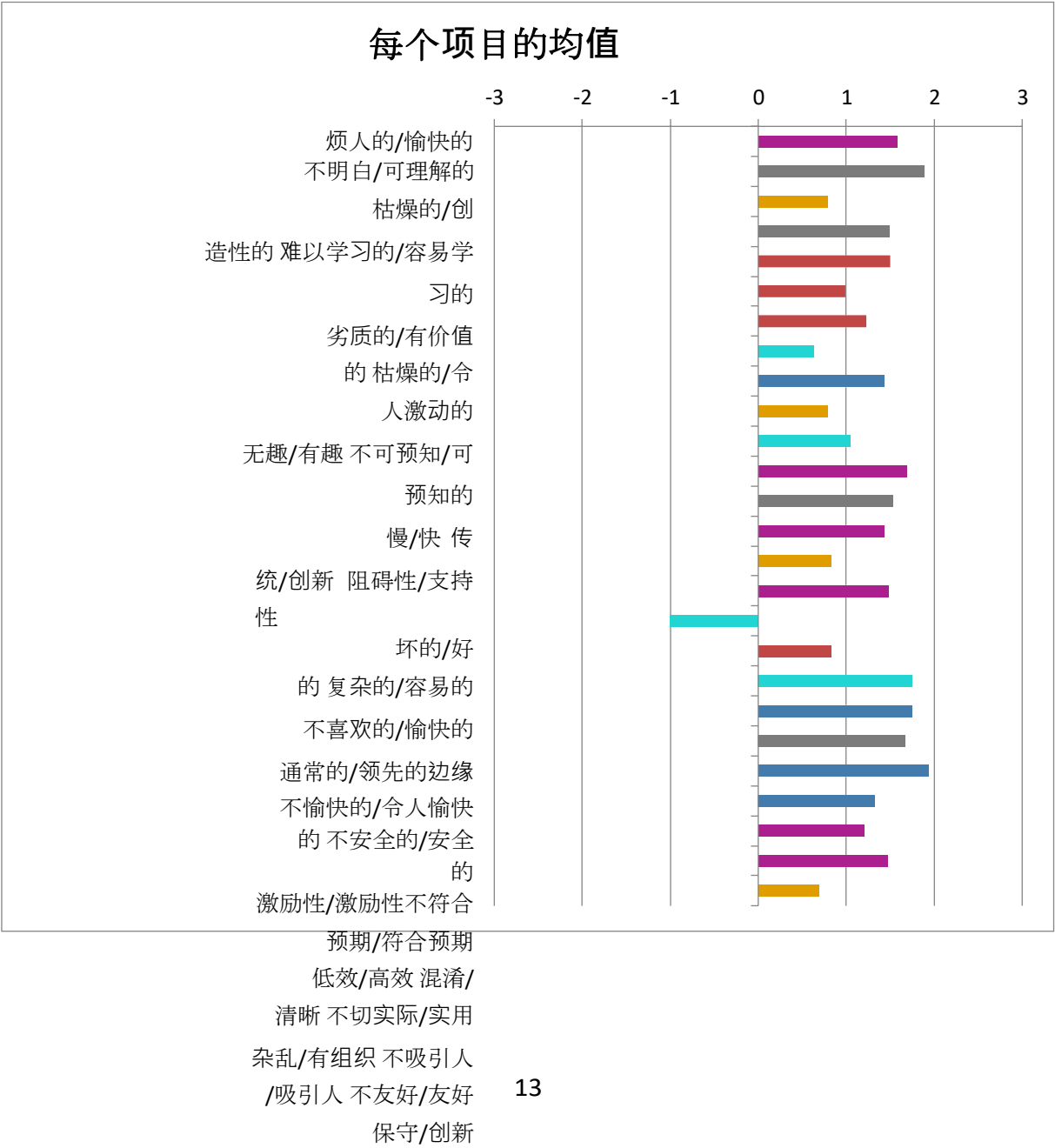


图5：在给定语境中被误解的项目的典型例子。

这个例子显示了UEQ在社交网络背景下的应用结果。很明显，项目17（不安全/安全）的平均值为负值，而这个量表的所有其他项目（黑条）的平均值为高度正值。这表明，在这种情况下，这个项目可能存在问题。

关于Alpha和样本量的影响的更多信息可以在（目前只有德语版）中找到。

*Schrepp, M. & Rummel, B., (2018). UX Fragebögen: Verwenden wir die richtigen Methoden? In: Dachzelt, R. & Weber, G. (Hrsg.), Mensch und Computer 2018 - Workshopband. Bonn: Gesellschaft für Informatik e.V. DOI: 10.18420/muc2018-ws16-0325.*

## 用于特殊应用场景的短版本

通常3-5分钟就可以完成包括一些人口统计学数据在内的UEQ的填写。因此，UEQ已经是一个相当有效的方法来捕捉用户对产品用户体验的意见。因此，一个自然的问题是，为什么还需要一个更短的版本？

在过去的几年中，我们收到了一些关于这种短版本的请求，一些用户甚至通过删除一些项目来创建他们自己的短版本（这不是一个推荐的做法）。因此，在实际应用中，似乎有一些用例，完整的UEQ太耗费时间了。

所有这些请求都来自三个不同的通用应用场景，其中只有很少的项目可以用来衡量用户体验。

1. 第一种情况是在用户离开网络商店或网络服务时收集数据。例如，用户刚刚在一家网店订购了一些东西，然后注销了。在按下注销按钮后，用户被要求填写一份关于商店的用户体验的简短问卷。在这种情况下，关键是要让用户感觉到填写问卷可以非常快地完成。显示一个包含所有26个问题的完整的UEQ将极大地限制愿意提供反馈的用户数量。
2. 在第二种情况下，关于用户体验的调查问卷应该被包含在已经存在的产品体验调查问卷中。一般来说，这样的问卷是在客户购买产品并使用一段时间后发出的。这样的调查问卷试图收集关于完整的产品体验的数据，例如客户为什么选择这个产品，产品的功能是否满足预期等等。因此，这类问卷往往相当冗长，因此很难增加一个完整的26项用户体验问卷。
3. 有时提到的第三种情况是实验环境，即被试需要在一次会议中判断几个产品或一个产品的几个变种的用户体验。在这种情况下，产品或产品变体以随机的顺序一个接一个地呈现给被试，被试必须填写一份关于每个产品的用户体验的调查问卷。在这种情况下，项目的数量必须保持在最低水平

。否则，参与者会感到紧张，答案的质量会受到影响。



为了满足这些要求，我们构建了一个短版的UEQ，它只包括8个项目。其中4个项目代表实用性质量（UEQ全文的项目11、13、20和21）和4个享乐性质量方面（UEQ全文的项目6、7、10和15）。因此，短版的UEQ并不提供对所有UEQ量表的测量。它包括实用性质量和享乐性质量两个量表。此外，还报告了一个总体量表（所有8个项目的平均值）。

你应该小心地解释整体规模。它不能被简单地解释为用户体验的整体KPI。原因是，对于被评估的产品来说，实用性质量和享乐性质量对用户的重要性可能是不一样的。对于某些产品来说，实用性质量可能比享乐性质量更重要，而对于其他产品来说，情况可能正好相反。因此，要谨慎解释总体价值。

由于短版的项目只是完整版的一个子集，所以短版可用于所有有完整UEQ翻译的语言。

为了能够快速处理这些项目，项目的负数总是在左边，正数总是在右边。首先显示的是实用性项目，然后是享乐性项目。

短篇UEQ的英文版本看起来是这样的。

阻碍性的	o o o o o o o o	支持性的
复杂的	o o o o o o o o	容易
效率低下	o o o o o o o o	高效
混乱的	o o o o o o o o	清楚
枯燥的	o o o o o o o o	激动人心的
没意思	o o o o o o o o	有趣的
常规的	o o o o o o o o	创造性
通常	o o o o o o o o	前沿

简易版UEQ的材料可以从[www.ueq-online.org](http://www.ueq-online.org)，在一个单独的软件包中下载。这个软件包包括一个包含所有语言项目的文件和一个用于数据分析的Excel工具。由于短版问卷的使用情况与纸笔问卷的使用情况不同，我们不提供短版问卷的标准说明，也不提供每种语言的PDF文档。

简短版本的UEQ-S仅用于特殊情况，不允许应用完整的UEQ。UEQ-S不允许测量详细的

用户体验质量 吸引力、效率、透彻性、可靠性、刺激性和新颖性，这些是UEQ报告的一部分。获得这些详细的数值对于解释结果和确定改进的领域是非常有用的。

因此，短版UEQ-S只允许对更高层次的元维度进行粗略测量。因此，我们的建议是，只在上述情况下使用短版UEQ-S。短版UEQ-S不应该取代标准情况下完整版的使用，例如在可用性测试之后。在这种情况下，小的收益

效率并不能弥补单一质量方面的详细信息的损失。

关于短版的构造细节，可以在以下文章中找到。

*Schrepp, Martin; Hinderks, Andreas; Thomaschewski, Jörg (2017): 用户体验调查问卷 (UEQ-S) 短版的设计和评估。In: IJIMAI 4 (6), 103-108. DOI: 10.9781/ijimai.2017.09.001.*

## KPI扩展

经理们喜欢KPI。他们将复杂的结构简化为一个单一的数值，从而给人一种印象，即事情可以通过测量和优化这个单一的KPI来控制 and 改善。然而，对于像用户体验这样结合了大量不同质量方面的结构，这并不容易做到。

UEQ给出了6个可以解释的刻度值。从调查问卷的设计来看，不可能将这些数值合并为一个单一的KPI。理论上可以证明的是，从效率、敏锐性和可靠性三个量表中计算出实用性质量的数值，从刺激性和新颖性中计算出享乐性质量的数值。

要从UEQ的结果中计算出一个单一的用户体验KPI，需要了解UEQ量表对用户体验总体印象的相对重要性。一种方法是增加一些问题，用简单的语言描述6个量表的内容，并询问参与者这些量表对他们的用户体验印象有多重要。然后，这些关于重要性的评分可以与量表手段相结合，计算出所需的用户体验KPI。一个质量方面对于整个用户体验印象的重要性显然取决于被评估的具体产品。因此，我们不可能提供任何关于量表重要性的一般数据。

如果你想从量表中计算出一个有意义的KPI，那么在UEQ中增加以下6个问题。

- **吸引力。** 产品看起来有吸引力，令人愉快，友好和令人愉快。
- **效率。** 我可以用该产品快速、高效、务实地完成我的任务。用户界面看起来很有条理。
- **透彻性。** 产品易懂、清晰、简单、易学。
- **可靠性。** 与产品的互动是可预测的，安全的，并符合我的期望。该产品支持我执行我的任务。
- **刺激性。** 使用该产品是有趣的、令人兴奋的和有动力的。
- **新颖性：** 产品具有创新性、发明性和创造性的设计。

这些问题可以用7级李克特量表来评分，终点是**完全不重要**（1）和**非常重要**（7）。附加问题应该放在26个UEQ项目的后面。通过重复最重要的积极术语，附加问题和相应的UEQ量表之间就建立了联系。

数据分析工具包含一个额外的工作表 *KPI\_Calculation*。在这个工作表中输入6个关于天平重要性的附加问题的观察值。然后，KPI将被自动计算。

额外的6个问题将在UEQ页面 ([www.ueq-online.org](http://www.ueq-online.org)) 的问卷调查 (PDF) 部分下载。请注意，这些问题目前还没有在UEQ的所有语言版本中出现。

有关这种方法的细节可以在以下文章中找到。

Hinderks, A., Schrepp, M., Domínguez Mayo, F.J., Escalona, M.J., Thomaschewski, J. (2019)。开发基于用户体验问卷的用户体验KPI。计算机标准与接口。  
<https://doi.org/10.1016/j.csi.2019.01.007>。

## 典型问题

### 我可以改变一些项目吗？

你不应该改变单一的项目，也不应该遗漏量表中的某些项目。如果你这样做，就很难解释你的结果，例如，不应该使用根据原始项目计算的基准值，因为答案没有可比性。

### 我需要所有的秤吗？

你可以不使用完整的量表，也就是说，从问卷中删除某个量表的所有项目。在某个量表显然不重要的情况下，这对缩短问卷调查是有意义的。

### 参与者需要多长时间来填写调查问卷？

应用UEQ不需要太多的努力。通常情况下，3-5分钟足以让参与者阅读说明并完成调查问卷。

### 我需要多少个数据？

你收集的数据越多，比例尺的平均值就越好、越稳定，因此你从这些数据中得出的结论就越准确。然而，为了得到可靠的结果，不可能给出你需要收集的最低数据数量。你需要多少数据还取决于参与问卷调查的用户同意程度（每个量表的标准差）。他们同意的程度越高，即项目答案的标准差越低，你需要的可靠结果的数据就越少。对于迄今为止评估的典型产品，大约20-30人已经给出了相当稳定的结果。

用于数据分析的Excel工具包含一个名为 *Sample\_Size* 的工作表。在这里，每个刻度的标准偏差被用来估计你需要多少数据才能使你的测量达到一定的精度（用置信区间的宽度来衡量）。很明显，精度取决于你想从数据中得出的结论。对于典型的产品评估，0.5的精度似乎已经足够了（见工作表的详细解释）。

## 如何向客户和管理层推销它？

如果你向你的管理层或其他利益相关者报告，传达UEQ量表的含义是很重要的。如果你评估一个财务软件或其他商业工具，如果你报告的是*刺激性*或*原创性*，对你的利益相关者来说可能会显得很奇怪。他们想要一个*原创的*会计系统吗？可能不是。

如果有必要，改变量表的名称，并明确解释每个量表的含义。在上面的例子中，你可能想把*Stimulation*改为*Fun of use*，把*Originality*改为*Interest*。使用适合你的利益相关者语言的术语。重要的是量表的语义，也就是说，如果你改变了一个量表的名称，要确保新的名称仍然涵盖了量表的含义。

## 是否有特殊产品类型的基准可用？

该基准允许将你的结果与其他产品的结果进行比较。然而，基准数据集包含相当不同类型的产品。因此，经常有人问，是否有特殊的基准数据集，只包含特定类型的产品。

由于UEQ基准的性质，显然需要一定数量的研究来创建一个有意义的基准。目前，我们只能为特殊产品类型提供两个专门的基准，即*商业软件*和*网页及网络服务*。这些基准和一般基准的边界显示如下。

一般基准（452个产品评价）。

类别	吸引力	敏锐性	效率	可靠性	刺激	原创性
优秀	1.86	2.03	1.90	1.70	1.70	1.61
良好	1.60	1.77	1.50	1.47	1.35	1.14
高于平均水平	1.19	1.25	1.06	1.15	1.01	0.75
低于平均水平	0.70	0.75	0.60	0.78	0.50	0.25

商业软件 (158个产品评价)

类别	吸引力	敏锐性	效率	可靠性	刺激	原创性
优秀	1.72	1.85	1.70	1.61	1.53	1.49
良好	1.50	1.48	1.38	1.37	1.34	1.03
高于平均水平	1.11	1.03	0.89	1.04	1.04	0.64
低于平均水平	0.70	0.58	0.51	0.71	0.71	0.24

网站和网络服务（85个产品评价）。

类别	吸引力	敏锐性	效率	可靠性	刺激	原创性
优秀	1.75	2.07	1.70	1.70	1.56	1.12
好的	1.41	1.84	1.43	1.53	1.10	0.87
高于平均水平	0.96	1.14	0.98	1.19	0.69	0.49
低于平均水平	0.44	0.65	0.50	0.81	0.07	-0.22

单元格中的数值意味着产品必须超过这个数值才能被列入相应的类别。因此，一个商业软件要被评为*优秀的透彻性*，必须超过1.85的数值，要被评为*良好*，必须在1.48和

1.85之间，如果低于0.58，则被评为不良。