

第11章 可用性测试

当你设计好一款产品，要如何去确认你的用户觉得这款产品好不好用？使用过程中会不会遇到问题？满不满意？产品设计过程中遇到一些颇感纠结的地方，不知道实际用户对产品会怎么理解和操作？一款产品开发出来后，想在投放市场前先检验一下这款产品是否靠谱？适不适合投放市场？针对以上这些情况，本章节介绍一种常用的产品检验方法——可用性测试。

11.1. 实验背景

11.1.1. 可用性测试的定义

可用性(usability)是衡量产品设计好坏的重要标准。国际标准 ISO 9241-11 将可用性定义为“特定产品在特定使用环境下，为特定用户用于特定用途时所具有的有效性(effectiveness)、效率(efficiency)和用户满意度(satisfaction)”。ISO 9241-210 进一步指出，更好的可用性会提高用户满意度、避免使用压力、增加可访问性和降低伤害风险。

可用性测试(Usability Testing)是通过观察和分析用户操作任务的行为，对产品可用性进行评估的一种方法。它可以在设计早期针对原型开展测试，也可在设计后期面向成品开展测试。可用性测试不仅有助于提升产品的用户体验，还可以提高用户对产品的接受度。多项研究发现，医护人员对电子健康记录系统的接受度和态度与系统的可用性密切相关(Jaspers et al., 2008; Huryk, 2010; Bleich & Slack, 2010)。目前，可用性测试已广泛应用于各类仪器设备、信息系统、软件网站、消费电子产品以及家用电器等的评价(Hafidz & Azizah, 2021)。

11.1.2. 可用性测试方法

可用性测试的方法有多种，但大致可分为观察法、调查法、专家评审法和模型法四类(董建明, 2007)。

(一) 观察法

观察法是一种观察用户与目标系统进行交互的方法，主要包括用户测试法和录像记录法两类。用户测试是指邀请用户参加可用性测试，这种测试一般安排在实验室里进行；录像记录法是用摄像机录下用户的使用过程，并通过事后分析找出与可用性有关的问题。

常用的用户测试法有实验室实验和出声思维(thinking aloud)两种方法。出声思维法要求用户在使用产品的同时将自己的主观感受与体验说出来,包括遇到的困难、问题、主观感觉等。使用该方法能从受测者对产品的评价中收集到大量信息。采用出声思维法需要注意招募具有代表性的用户、制定具有代表性的任务与操作、以及自始至终耐心听用户讲述。

(二) 用户调查法

用户调查法包括问卷法和用户访谈法。问卷法所收集的数据可直接用来进行统计分析,因此对问卷本身需作较为全面的考虑和周密的设计,以保证研究者能收集到可信的数据,且涵盖需考虑的所有可用性问题。问卷法的实施大致可分为明确调研目的、确定调研内容、发放调研问卷以及分析调研结果四个阶段。常见的可用性问卷包括软件可用性测量问卷(software usability measurement inventory, SUMI)、计算机系统可用性问卷(computer system usability questionnaire, CSUQ)、用户交互满意度问卷(questionnaire for user interaction satisfaction, QUIS)、普渡大学可用性测试检查表(Purdue Usability Testing Questionnaire, 见附录 12.6.1)等。

用户访谈一般安排在用户使用产品后进行,主要目的是了解用户的主观感受。访谈有开放式访谈、结构化访谈、半结构化访谈三种类型。具体采用何种访谈技术则取决于评估目标、待解决的问题和选用的评估类型。例如,若目标是为了了解用户对交互设计的反应,非正式的开放式访谈通常是较好的选择;若目标是从用户中搜集关于产品某些特定特征的反馈意见,则采用目标和问题更为具体的结构化访谈可能更为合适。感兴趣的读者可进一步阅读《洞察人心 用户访谈成功的秘密》(Portugal, 2023)一书,对访谈法取得更全面的了解。

(三) 专家评审法

专家评审法是由可用性专家来评估目标人机系统可用性的一种方法。专家评审法大致可分为认知走查(cognitive walkthrough)和启发式评估(heuristic evaluation)两类(Jadhav, 2013)两种方法。认知走查是评估人员或专家从用户学习使用系统的角度来评估系统的可用性。该方法主要针对新用户使用指定产品时可能遇到的问题。所谓走查,就是一步一步全面检查使用、执行的整个过程,从中找出可用性问题。

启发式评估法是由 Jakob Nielsen 等人开发的一种非正式的可用性检查方法。实施启发式评估时，由 3~5 名评估人员或专家使用一组称为“启发式原则”的可用性规则作为指导，评定用户界面各个元素(如对话框、菜单、在线帮助功能等)是否符合这些原则。进行启发式评估时，评估人员须运用与目标用户相同的视角和目标，模拟典型用户使用产品的情形，从中找出潜在的可用性问题。启发式评估不需要用户参与，也不需要特殊设备，这种评估方法成本较低并且快捷。

(四) 模型法

模型法通常是用数学模型来量化人机交互的过程。例如，当需要对一个手机 APP 采用模型法进行可用性测试时，需首先将 APP 分解为一个个不同的功能，每个功能又被细分为多个子功能，然后测出完成每个子功能所需要的时间以及最终掌握整个 APP 功能所需的时间。模型法特别适合于无法直接通过用户进行可用性测试的情形。GOMS(Goals, Operators, Methods, Selections)模型是模型法的典型代表之一。

近年来，WOZ(Wizard of Oz)法在人机交互和人机界面设计领域也得到了比较广泛的应用(Kelley, 1984)。在中文文献中，“Wizard of Oz”大多被译为“绿野仙踪”。WOZ 法的特点是建立起一个模拟的交互式系统或产品的使用场景，但实际上某些功能是由专业人员(Wizard, “魔法师”或“巫师”)来完成的，而非真正运行中的自动化系统。这一方法的名称来自于美国作家弗兰克·鲍姆创作的童话作品《绿野仙踪》，通过人在幕后操作带来前台魔法般使用的演出效果。这一方法的出现是因为在设计初期或者技术尚未完全实现之时，需要评估用户体验和系统反应的可行性。WOZ 法在新产品设计、用户体验测试以及探索新型交互方式等方面都能发挥很好的作用。这种方法因其具有较高的灵活性和有效性，而持续受到关注和被广泛应用。特别是在 AI 时代，WOZ 方法可用于帮助用户研究人员模拟 AI 驱动下的产品或设计的使用场景，而得到了越来越多研究者的关注(综述见 Riek, 2012)。具体 WOZ 的操作步骤可参见附录 12.6.4.WOZ 方法介绍。

11.2. 实验目的

本章主要介绍如何在可用性实验室中综合运用可用性测试的几种常用方法，有效开展可用性测试。实验目的是：

(一) 了解可用性测试的概念与方法

了解和掌握可用性测试的概念，熟悉可用性测试常用方法。

(二) 掌握可用性测试的流程

通过实践，熟悉并掌握可用性测试的一般流程，学习如何撰写可用性报告

11.3. 实验方法

11.3.1. 实验设备

专业的可用性测试一般是在可用性测试实验室内进行。可用性测试实验室一般由测试室和观察室组成，两者间有一面隔音的单向玻璃，允许施测人员通过单向玻璃观察受测者的操作而不影响受测者的行为。测试室中一般会配备行为采集系统(视频和音频采集系统)，用于记录受试者在实验过程中的动作、表情和声音。测试室中还会安装用于传送测试人员指令的传声装置。观察室中安装可用性测试专用设备、用户行为观察分析系统、行为控制设备、摄像机等。图 11.1 是典型的可用性测试实验室的示意图。值得一提的是，在受到外界环境(如疫情)、任务性质(如任务需要在用户自己的设备上完成)等因素的限制而无法在实验室内实施可用性测试时，根据需要也可以开展在线可用性测试(Usability Testing 101)。

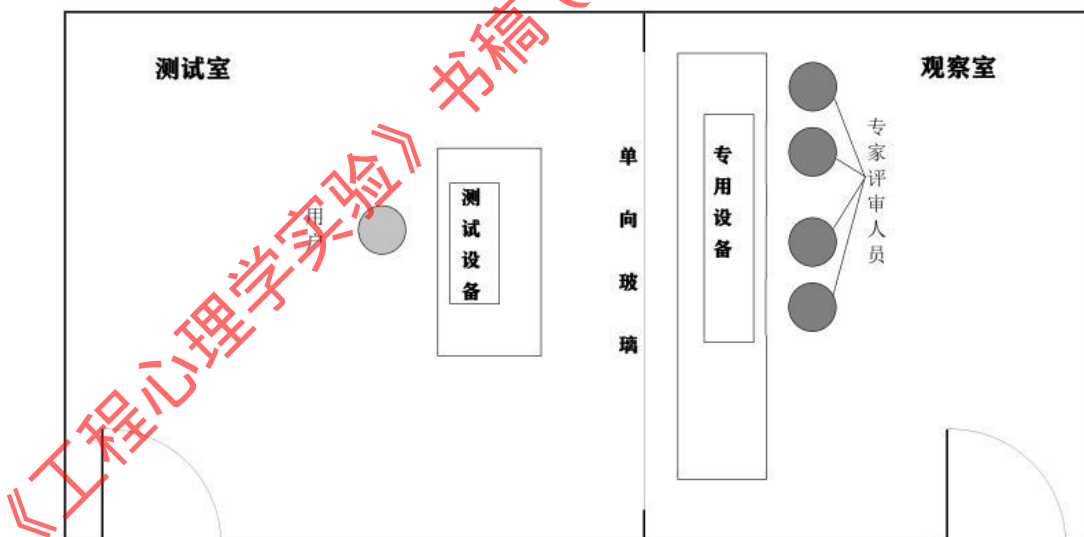


图 11.1 可用性实验室图例

11.3.2. 实验设计与流程

可用性测试一般分为前期准备、正式测试、分析报告、后期跟进四个步骤。

步骤一、前期准备

前期准备工作主要包括制定任务、准备测试辅助材料和用户招募等。

制定任务 制定可用性测试任务是可用性测试前期准备工作的难点之一。在这一阶段，测试人员一般需考虑以下三方面的有关事项：(1) 测试的目标，(2) 测试内容的精准呈现，(3) 拟从测试中获取的数据与结论。制定可用性测试任务时可参考以下 6 项原则：

1. 为测试任务设定合理的情景。这有助于受测用户更好地理解测试内容，更快地进入测试工作。任务情景的描述可参照以下格式：用户在一定情况下(通常应为真实情景，这样才能观察到用户在自然场景中暴露出的问题)，想要完成什么样的任务(任务需要具有合适的颗粒度，既使用户能完成对测试功能的操作，又不能对其过度引导)。

2. 任务设置要周密、全面，确保完全覆盖要测试的各项功能。

3. 任务执行顺序要符合产品正常使用的流程。

4. 告知受测用户要达到的目标(即要实现的任务)而非仅仅提示操作步骤。明确目标任务能保证受测用户在不偏离主题的范围进行测试。在测试过程中，测试人员不可以向受测用户提示操作步骤或施加干扰，以免影响测试的科学性，进而影响问题的发现与判断。

5. 实验任务要易于理解，无歧义。

6. 任务数量不宜过多，并且完成任务所需的时间一般应控制在 30 至 60 分钟以内。测试时间过长会使受测用户因疲劳而失去耐心。

下表 11.1 结合上述原则采用具体事例对做任务描述时需要注意的事项作了说明。

表 11.1 关于用户使用携程上购买机票的功能的任务描述

任务表达	评价
在携程 APP 里，使用“买机票”功能	该描述太具体、并且向用户提示了操作步骤。用户若按照指示操作，可能会遗漏自己独立操作过程中出现的问题
请用携程 APP 买一张机票	缺少真实情景的铺垫，用户仅完成操作任务，无法暴露一些真实情景下可能出现的问题
爸妈让我去买飞机票	描述过于宽泛，遗漏了测试功能。用户

	可能不用测试产品去完成任务，或用了产品但遗漏了想要测试的功能。
本周末准备和父母一起去三亚旅游，“我”负责在携程上预订三人往返三亚的机票	较为明确的表述

测试辅助材料 测试人员需根据可用性测试的流程事先准备好各个环节中要用的材料，如测试前的访谈提纲、主试脚本(附录 12.6.2)、任务检查单、观察者指南(附录 12.6.3)等。访谈提纲一般需涉及以下内容：(1) 受访用户的基本信息，如年龄、职业、兴趣爱好、生活工作状态、手机使用、互联网使用情况等，这些基本信息的收集与统计有利于后续对参与测试的用户进行分类和数据分析。(2) 相关产品的使用经验和习惯，以及用户是否对产品有个人的想法和功能需求。如果对已经投放市场的产品进行更新换代，用户是否会保留一些自己当前的用法和习惯；如果用户未使用过待测产品，需要了解用户是否有相关产品的使用经验。

招募用户 在开展可用性测试前，要明确招募用户的数量。可根据具体的测试内容和开展测试的客观条件，确定需要招募的用户数量。Nielsen 研究表明，参加可用性测试的用户达到 4 人时就可发现 75% 的可用性问题，有 5 人参加的测试，可以发现 85% 的产品可用性问题。所以，招募用户的数量一般应在 4-5 人以上(Nielsen,1993)。

招募用户时需选取产品潜在的目标用户群或具有代表性的用户群。因此，招募用户时要考虑如下两个因素：第一，是选用从未接触过产品的新手、还是熟练使用者？或是两者按比例选取？第二，用户的性别、年龄、教育水平等因素。可设计一个表格来统计用户群体的人口学特征，如表 11.2。

表 11.2 用户表格设计样例

姓名	年龄	性别	教育水平	职业	计算机经验	产品经验

步骤二 正式测试

典型的可用性测试通常包括测试前访谈、测试、测后反馈三个环节。有的可用性测试涉及到一些与伦理有关的问题，测试前还需要与用户签订关于录音录像

许可、个人隐私保护、知情同意和保密等方面的协议书。

在对产品正式测试前，测试人员按以下步骤做好实验设备检查和其他各项准备工作：

1. 准备好所有需要打印的材料，如主试脚本、任务检查单、观察指南等。
2. 检查实验室内各项设备的状况，如计算机是否符合测试需求，摄像头位置是否正确(摄像头需确认可清晰拍摄计算机或其他测试设备、受测者面部等)，传声装置位置是否正确。
3. 检查观察室内各项设备的状况，如摄像机软件是否正常运行同步的画面和声音是否清晰、不卡顿，传声装置是否正常等。

访谈 访谈的目的是为了充分了解用户。在访谈过程中，需要让用户了解访谈的目的，打消思想顾虑，解除防御心理、畅所欲言。建议测试人员可首先询问用户几个简单的问题，让他放松，并表明将仔细倾听。然后，测试人员向用户正式提问，获取所需要的信息。与用户的访谈可采用 5W1H (Who、Where、When、What、Why、How) 提问法，其中的 Who 指的是产品使用对象、Where 指的是使用场景、When 指的是使用时间、Why 指的是使用功能、How 指的是选择使用或拒绝使用的原因以及如何使用。访谈中进行的对话要尽量明确、具体，避免使用“一般”“大多”等概括性词汇；所作的描述要尽量口语化，避免过于专业的词汇。在采访用户了解他对当前产品或竞品的个人使用经验时，可采用图 11.2 所示的流程。

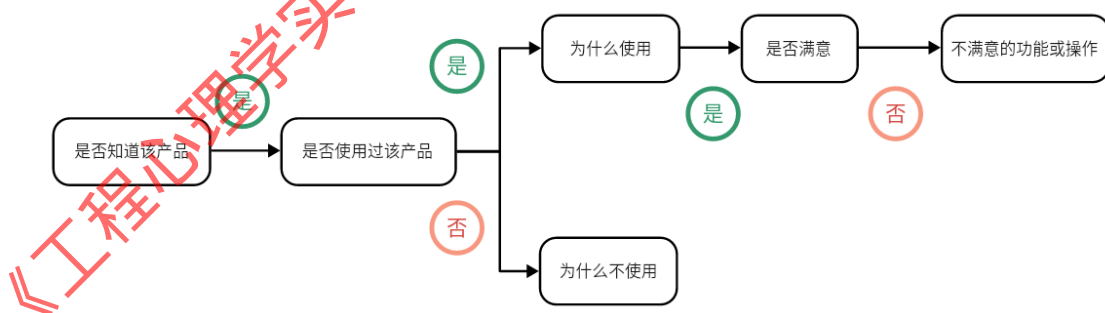


图 11.2 用户访谈一般流程

测试 按照测试方案让用户逐项完成指定的操作任务，直至完成整个测试。在任务执行前，测试人员先向被试详细介绍所要测试的任务，包括任务名称和内容等。用户开始执行任务后，尽量不要打断他们，让他们把注意力集中在完成当前的任务上，并且一边执行任务，一边大声说出自己的感受与评价。在下列情况

下，可考虑进入下一项任务：(1) 用户完成了任务；(2) 用户情绪低落，时间不多而测试人员对后续任务更感兴趣；(3) 继续下去并不能得到有价值的信息；(4) 在用户完成测试任务后可继续询问他是否还有其他方面的问题。

在测试过程中要注意做好两类数据的收集工作：一类为客观定量的数据，如完成任务需要的时间，操作任务过程中犯错误的次数等；另一类是主观评价的定性的数据，如产品满意度、接受度。用户对产品的主观评价可以在实验结束后采用访谈、问卷调查等方法来完成；也可在实验中采用出声思维的方法来完成。

在测试过程中，测试人员应与受测用户保持及时的沟通。可以由一位测试人员与用户进行沟通，另一位测试人员负责观察和记录。除了记录来自用户的信息外，也要记录负责沟通的测试人员在访谈过程中是否存在问题。在对用户进行产品功能测试过程中，研究人员可根据前期访谈中获取的内容作为观察重点，并尽量关注用户操作过程中的一些细节，从中找到存在的问题以及需要受测用户进一步提供哪些方面的信息，如询问用户你是怎么找到这一功能/界面的，你觉得这一处的设计是什么意思，为什么采取这样的操作，对现在这种操作方式有什么样的感受，在整个操作过程中有什么感受等。

测试过程中与用户的沟通需注意以下事项：

1. 避免过早告诉用户要测试的产品、产品功能的定位、使用方法。要让用户自行去探索产品的定位、使用方法，观察用户是否理解设计目标。
2. 避免直接告诉用户如何完成任务。观察用户是如何探索产品的，并在恰当的时间追问用户当下的感受。
3. 若用户跳过了要测试的功能，测试人员应避免过早提问，不要脱离情景去发问。测试人员需耐心等待用户自我操作并观察用户的操作情景，最后询问跳过或不使用该功能的原因。
4. 用户和研究员对任务完成的理解可能不一致，用户即使完成某项操作也有可能并不理解设计的规则和结果。此时，测试人员需追问用户对操作或产品功能的理解程度。
5. 避免使用带有“诱导性”的词汇来提问用户的感受。如“你是因为××设计才不用这个功能吗？”要多倾听用户的想法，及时追问用户不满意的原因。
6. 避免表达测试人员的个人态度或过多评价用户的行为和想法。

7. 避免让用户直接给出解决方案。测试人员需追问清楚用户的意图和困难，在测试完成后再据此思考改进方案。

8. 避免让用户脱离其产品使用的角度来总结感受；要基于用户使用的角度来提问，推测出用户想要的结果。

测后反馈 在用户完成某一功能的测试后，可通过再次访问的方式收集用户的反馈(如对测试进行简单打分)。以此来深入了解用户的想法，保证测试数据和结果更具时效性和真实性。可用性测试有时需要在最后阶段对用户进行满意度问卷调查。问卷可采用检查单或开放性等问题等形式。

可用性测试的所有环节完成后，对用户表示感谢，然后询问用户是否还有疑问，并予以及时回应。最后，向用户道别，并开始着手准备下一场测试。

步骤三 分析报告

完成可用性测试后，测试人员对用户测试的录音录像、问卷调查、用户访谈或出声思维等环节中发现的问题进行分析和总结，例如，任务效率分析可采用类似于表 11.3 和 12.4 的表格来完成。测试人员将在可行性测试中收集到的用户反馈和各类相关信息提炼为最终需要解决的可行性问题，并对问题作出准确描述。可根据可用性测试的三个标准(有效性、效率、满意度)，对由测试找出的问题进行分辨，挑选出符合可用性研究范围内的问题。值得一提的是，用户的个人喜好与个人需求会影响对产品的满意度，但它们只是属于个人偏好问题，不属于操作存在障碍或设计不合理、有缺陷，故不应将其归于可用性问题。

表 11.3 单个任务效率分析表示例

被试号	独立完成任务百分比	辅助完成任务百分比	任务完成时间	协助类型	备注

表 11.4 整体效率分析表示例

被试号	整体独立完成任务百分比	整体辅助完成任务百分比	整体任务完成时间	整体协助类型	备注

对可用性问题进行描述时应注意以下事项: (1) 从用户角度说明遇到的问题，而非从产品角度说明存在的问题。(2) 还原问题发生的情景，而非只是阐明问题。

(3) 客观真实地描述用户目标、行为、想法。(4) 可根据问题的严重性区分问题解决的优先级。可根据问题发生是否属于核心流程和是否会影响任务完成这两个方面来对问题分级。非核心流程最终完成的任务属于轻微问题，非核心流程最终无法完成的属于中等问题，核心流程最终完成的属于严重级别问题，而核心流程最终无法完成的属于灾难级别问题。后两种问题需要优先解决。基于上述注意事项，一问题描述可参照以下格式：用户在某个问题场景下想要完成某件事，发生了某个操作行为，导致了某个问题结果。

最后，测试人员将系统存在的各种可用性問題加以汇总和撰写可用性测试报告，并将可用性测试报告提交产品设计部门作必要的修改。建议最终挑出不超过 10 个最严重的问题。将问题按其严重程度排序，然后按照问题严重性由高到低的次序讨论解决方案。

步骤四 后期跟进

当设计师根据可用性测试的反馈信息完成产品的改进与优化后，可用性测试人员可对新的产品版本进行测试，对比改进前后的测试结果，并检查改进、优化的效果。

11.4. 实践

以小组为单位(建议每组 3-4 人)，设计一个可用性测试方案，并进行实地测试。测试任务由各组自定，可以是网页测试，手机客户端测试或 APP 测试等。

根据可用性测试的结果，撰写可用性测试报告。报告需包括测试产品介绍、测试方法和测试流程、数据分析、可用性问题分析、解决方案与建议等内容。

11.5. 思考题

1. 可用性测试与用户体验有什么样的关系？
2. AB test 是用户研究常用的方法之一，该方法是否可用于可用性测试？为什么？

11.6. 附录

11.6.1. 普渡大学可用性测试检查表(<https://garyperelman.com/question/question.cgi?form=PUTQ>,

Lin, Choong, & Salvendy, 1997)

使用说明：本调查表共有 100 题，请回答每一个问题，每一题按照如下三个步骤回答：

(a) 请评估每一个问题是否适用于所评定的系统，如果不适用，跳到下一题，如果适用，请继续回答下面两个问题。

(b) 请评价该问题对所评定系统的重要程度(1 最不重要的，3 是最重要的)。

(c) 评价系统在该问题上的表现(1 非常糟糕，7 非常好)；如果不存在，请选择不存在这一项。

一. 兼容性	
1.光标的控制是否符合光标的移动?	2.用户控制的结果是否符合用户的期望?
3.所提供的控制是否符合用户的技能水平?	4.界面的编码(如颜色、形状等)是否为用户所熟悉?
5.用词是否为用户所熟悉?	
二. 一致性	
6.界面颜色的编码是否符合常规?	7.编码是否在不同的显示及菜单上都保持一致?
8.光标的位置是否一致?	9.显示的格式是否一致?
10.反馈信息是否一致?	11.数据字段的格式是否一致?
12.标号的格式是否一致?	13.标号的位置是否一致?
14.标号本身是否一致?	15.显示的方向是否一致?(漫游或卷动)
16.系统要求的用户动作是否一致?	17.在不同的显示中用词是否一致?
18.数据显示和数据输入的要求是否一致?	19.数据显示是否符合用户的常规?
20.图形数据的符号是否符号标准?	21.菜单的用词和命令语言是否一致?
22.用词是否符合用户指导的原则?	
三. 灵活性	
23.是否可以使用命令语言而绕过菜单的选择?	24.系统是否有直接操作的功能?
25.数据输入的设计是否灵活?	26.用户是否可以灵活地控制显示?
27.系统是否提供了灵活的流程控制?	28.系统是否提供了灵活的用户指导?
29.菜单选项是否前后相关?	30.用户是否可以根据需要来命名显示和界面单元?
31.系统是否为不同的用户提供了好的训练?	32.系统是否可以自己改变视窗?
33.用户是否可以自己命名系统命令?	34.系统是否允许用户选择需要显示的数据?
35.系统是否可以提供用户指定的视窗	36.为了扩展显示功能，系统是否提供了放大的功能?

四.可学习性	
39.命令语言是否有层次?	40.菜单的分组是否合理?
41.菜单的顺序是否合理?	42.命令的名字是否有意义?
43.系统是否提供了无惩罚的学习?	
五.极少化的用户动作	
44.系统是否为相关的数据提供了组合输入的功能?	45.必要的数据是否只需要输入一次?
46.系统是否提供了默认值?	47.视窗知觉的切换是否容易?
48.系统是否为经常使用的控制提供了功能键?	49.系统是否有全局搜索和替代的功能?
50.菜单的选择是否可以使用点击的功能?(主要的流程控制方法)	51.菜单的选择是否可以使用键入的功能?(辅助的控制方法)
52.系统是否要求极少的光标定位?	53.在选择菜单时,系统是否要求极少的步骤?
54.系统是否要求极少的用户控制动作?	55.为了退到更高级菜单中,系统是否只需要一个简单的键入动作?
56.为了退到一般的菜单中,系统是否只需要一个简单的键入动作?	
六.极小化的记忆负担	
57.系统是否使用了缩写?	58.系统是否为输入分层次的数据提供了帮助?
59.指导信息是否总是可以得到?	60.系统是否为序列的选择提供了分层次的菜单?
61.被选的数据是否有突出显示?	62.系统是否为命令提供了索引?
63.系统是否为数据提供了索引?	64.系统是否提示在菜单架构中的当前位置?
65.数据是否保持简短?	66.为选择菜单使用的字母代码是否经过认真的设计?
67.是否将长的数据分成不同的部分?	68.先前的答案是否可以简便的再利用?
69.字母大小写是否等同?	70.系统是否使用短的代码而不使用长的代码?
71.图符是否有辅助性的字符标号?	
七.知觉的有限性	
72.系统是否为不同的数据类别提供不同的编号?	73.缩写是否清晰而互补相同?
74.光标是否不同?	75.界面单元是否清晰?

76.用户指导的格式是否清晰?	77.命令是否有清晰的定义?
78.命令的拼写是否清晰?	79.系统是否使用了易于分辨的颜色?
80.目前活动的窗口是否有清楚的标识?	81.为了直接比较,数据是否成对地摆在一起?
82.是否限制语音信息使用的数量?	83.系统是否提供了一系列相关信息?
84.菜单是否和其它的显示信息有明显的区别?	85.颜色的编码是否多余?
86.系统是否提供了视觉上清晰可辨的数据或文字?	87.不同组的信息是否明显分开?
88.屏幕的密度是否合理?	
八.用户指导	
89.系统反馈的错误信息是否有用?	90.系统是否提供了“取消”的功能?
91.错误的输入是否被显示出来?	92.系统是否提供了明确的纠错方法?
93.系统是否为控件输入提供了反馈?	94.是否提供了“帮助”?
95.一个过程的结束是否标志清楚?	96.是否对重复的错误有提示?
97.错误信息是否具有建设性并提供有用的信息?	98.系统是否提供了“重新开始”的功能?
99.系统是否提供了“撤销”的功能?	100.用户是否启动流程控制?

11.6.2. 网页可用性测试主试脚本示例

以下是供主试参考的网页可用性测试脚本示例,是基于《妙手回春:网站可用性测试及优化指南》(Krug, 2014)修订的。

欢迎阶段

您好, ____。我是今天的主试,叫____,今天将由我来引导您完成测试。

Web 浏览器目前显示的是____网页。在我们开始测试之前,需要让您了解一些信息。我将逐字逐句为您宣读,以免遗漏。

您可能很清楚我们请您来这里的目的,但请允许我在这里再简单地重复一遍。我们请您来使用我们开发的网站,以确定它是否按预期的那样工作。测试将持续大约 30 分钟。在这里首先要澄清的一点是,我们要测试的是网站而不是您。您在这里不会做错任何事情,您无需为自己的可能犯错担心。

当您使用网站时,请您尽可能说出来您看到的、想做的以及怎么想的,这将给我们提供极大的帮助。请不用担心我们的感受。我们做这项测试只是为了改善网站的设计,因此需要知道您的真实反应。

如果您在测试期间有任何问题,可以随时询问。但我无法立即回答,因为我们想知道在没有人为你提供帮忙的情况下你将如何做。如果测试结束后您还有问题要问,我将尽最大努力做出回答。另外,您可以在任何时间休息,请随时告知我们。

<p>您可能注意到了这里的麦克风。在您允许的情况下，我们将把屏幕上发生的情况以及我们之间的谈话录下来。录像只是用来帮助我们确定改进网站的方法，而不会被与该项目无关的任何人看到。另外，有几位 Web 设计小组的成员在玻璃屏幕后的另一个房间观看这次测试。</p> <p>(如果准备了录像许可表)如果您愿意，我想让您在这个许可表上签字。它只是指出录像是得到您许可的，而录像只会被该项目相关的人看到。</p> <p>(如果准备了保密协议)我们还事先给您发了一份保密协议，上面规定您不能与任何人讨论今天看到的東西，因为我们还没有发布它。您带来了那份保密协议吗？请问是否还有什么问题吗？</p>
<p>提问环节(可选择性的询问)</p>
<p>在使用网站前，我想问几个简单的问题。首先，您从事哪种职业？每天都做些什么？您每周上网大概有多少小时？这包括上班時間以及在家里浏览网页和收发邮件。收发邮件和浏览网页各占多大比例？浏览网页时都访问什么样的网站？有非常喜欢的网站吗？好，就这样。</p> <p>问题问完后，开始观看网站。</p>
<p>主页浏览(确保该过程最多不超过三四分钟)</p>
<p>单击指向网站主页。</p> <p>首先，请您观看这个页面，并告诉我您是如何理解它的？什么地方最吸引您？您认为这是什么公司的网站？您能在该网站上做什么？该网站是做什么的？</p> <p>只需要观看并作出简单描述即可。如果您想要使页面向下滚动，可以这么做，但是不要点击页面上的任何地方。</p>
<p>任务执行</p> <p>下面我将让您尝试完成一些具体的任务。我将宣读每项任务，并给您提供打印好的任务情景描述。</p> <p>注意，执行这些任务时请不要使用搜索功能，这样能让我们更深入地了解网站是不是像预期的那样运行。同样，请在执行任务的时候尽可能大声说出你的感觉与体验，这将对我们有很大帮助。</p>
<p>问题探讨：</p>
<p>如果您不介意，请给我一分钟，让我确认小组成员是否有其它问题需要您回答。</p> <p>与观察室联系，询问观察者是否有其他问题。向用户提出观察者的问题，然后根据需要进行测试。</p> <p>测试结束后，询问观察者对这次试验是否存在疑问。</p> <p>向用户表示感谢。</p>

11.6.3. 网页可用性测试观察者指南

<p>可用性测试的观察人员需事先了解整个测试的节奏，如今天总共有几个测试项目，每个项目的测</p>

<p>试持续几分钟，每两场测试间休息几分钟；今天的测试情境以及是否准备了备用的测试情境。</p> <p>为了从这些测试得到最大的收获，观察者需做到如下几点。</p>
<p>记笔记</p> <p>随时准备记录有用的信息，尤其是遇到感到迷惑或无法完成任务的情况。如果有问题想问主试，将问题写在纸上。每场测试快结束时，主试应询问观察人员是否还有其他问题。</p>
<p>在每场测试结束后制作检查单</p> <p>在两场测试之间的休息时间，记下在测试期间注意到的 3-4 个最严重的问题。</p>
<p>参与总结会讨论</p> <p>核对与比较笔记，以决定要处理哪些可用性问题。</p>
<p>尽量不要打扰他人</p> <p>观看测试必须高度集中注意力。尽可能不谈论与测试无关的内容，不要在观察室进行其它讨论或接听电话，不可将观察室当作电影院，谈话声音不可太大，谈话时间也不可太长。</p>

11.6.4. WOZ 操作步骤介绍

WOZ 按照以下步骤操作：

确定测试目标：明确要测试的功能、交互或用户场景。

1. 设计场景：创建一个模拟的环境或使用案例，确保它能够模拟真实用户可能遇到的情况。
2. 指定“巫师”角色：选择能够熟练操作系统或产品的人员，他们将扮演“巫师”的角色，负责模拟系统的反应或提供信息。
3. 模拟交互：让参与测试的用户与“系统”交互，他们可能会认为是自动化的系统，但实际上是“巫师”在背后进行操作。
4. 记录和分析反馈：记录用户的行为、反馈和体验，这可以通过观察、访谈、问卷调查等方式进行。
5. 调整和改进：根据收集到的数据和反馈，对产品或系统进行改进。

在实践过程中，研究人员需注意以下事项：(1) 保持逼真性：让模拟尽可能接近真实系统的操作和响应，以获得更真实的用户反馈。(2) 灵活应对：“巫师”需要灵活应对用户的行为和需求，模拟出自然的系统反应，同时不暴露实际的技术限制。(3) 数据收集：不仅要记录用户的行为，还需要记录他们的情感、反馈和认知过程，以获得全面的信息。

11.7. 参考资料

- 董建明(2007). *以用户为中心的设计和评估*(2 版), 清华大学出版社
- Bleich, H. L., & Slack, W. V. (2010). Reflections on electronic medical records: when doctors will use them and when they will not. *International Journal of Medical Informatics*, 79(1), 1 – 4.
- DIS, I. (2009). 9241 – 210: 2010. *Ergonomics of human system interaction-Part, 210*.
- Firdaus, H., & Zakiah, A. (2021). Implementation of usability testing methods to measure the usability aspect of management information system mobile application (case study sukamiskin correctional institution). *International Journal of Modern Education & Computer Science*, 13(5), 58 – 67.
- Huryk, L. A. (2010). Factors influencing nurses' attitudes towards healthcare information technology. *Journal of Nursing Management*, 18(5), 606 – 612.
- ISO9241 – 11.(1998), *Ergonomic requirements for office work with visual display terminals (VDT's) – Part 11:Guidance on usability*. International Standards Organization.
- Jadhav, D., Bhutkar, G., & Mehta, V. (2013, September). Usability evaluation of messenger applications for Android phones using cognitive walkthrough. In *Proceedings of the 11th Asia pacific conference on computer human interaction* (pp. 9 – 18).
- Jaspers, M. W., Peute, L. W., Lauteslager, A., & Bakker, P. J. (2008). Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system. *Studies in Health Technology and Informatics*, 136, 303 – 308.
- Kelley, J. F. (1984). “An iterative design methodology for user-friendly natural language office information applications.” *ACM Transactions on Information Systems (TOIS)*, 2(1), 26 – 41.
- Krug, S. (2009). *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems*. New Riders.
- Krug, S. (2014). *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems (2nd edition)*. New Riders.

- Lin, H.X. Choong, Y.-Y., and Salvendy, G. (1997) A proposed index of usability: A method for comparing the relative usability of different software systems. *Behaviour & Information Technology*, 16:4/5, 267 – 278.
- Nielsen, J., & Landauer, T. K. (1993, May). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 206 – 213).
- Portugal, S. (2023). Interviewing users: How to uncover compelling insights (2nd edition). *Rosenfeld Media*
- Riek, L. D. (2012). Wizard of oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1), 119 – 136.
- Usability Testing 101, <https://www.nngroup.com/articles/usability-testing-101/>

《工程心理学实验》书稿（尚未定稿，仅供教学使用）