



中华人民共和国国家标准

GB/T 25000.62—2014/ISO/IEC 25062:2006

软件工程 软件产品质量要求与 评价(SQuaRE) 易用性测试报告 行业通用格式(CIF)

Software engineering—Software product quality requirements and evaluation
(SQuaRE)—Common industry format (CIF) for usability test reports

(ISO/IEC 25062:2006, IDT)

2014-09-03 发布

2015-02-01 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

前言 I

引言 II

1 范围 1

2 符合性 1

3 规范性引用文件 1

4 术语和定义 2

5 报告格式 3

 5.1 标题页 3

 5.2 执行摘要 3

 5.3 引言 4

 5.4 方法 4

 5.5 结果 9

 5.6 附件 10

附录 A (资料性附录) 检查表 11

附录 B (资料性附录) 术语 14

附录 C (资料性附录) 报告模板 17

参考文献 23

前 言

本部分是 GB/T 25000《软件工程 软件产品质量要求与评价(SQuaRE)》的第 62 部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

本部分使用翻译法等同采用 ISO/IEC 25062:2006《软件工程 软件产品质量要求与评价(SQuaRE) 易用性测试报告行业通用格式(CIF)》。

本部分由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本部分起草单位:广西达译商务服务有限公司、山东省计算中心、中国电子技术标准化研究院、北京凯思昊鹏软件工程有限公司、北京邮电大学、北京中科院软件中心有限公司。

本部分主要起草人:邓姿娴、施贵基、丁锦春、马骏、邹丰义、杨美红、李刚、张倪、顾玉良、奉旭辉、刘建毅、高翔、王枳、袁玉宇、郭新伟、周鸣乐、韩庆良、顾卫东、董火民、韩红强。

引言

易用性是预测软件能否成功部署的关键因素。软件厂商在产品开发的阶段都进行易用性测试。一些企业在作出软件购买决策前也测试软件的易用性。测试通常包含以下内容：(1)软件目标用户中具有代表性的受测人；(2)具有代表性的任务；(3)效率、有效性和主观满意度的测度。包含以上内容的测试被称为总结性测试，即，其结果是具有统计意义的测度趋势（如平均数或中间值）和可变性（如标准差）。易用性测试报告行业通用格式（CIF）目的在于辅助易用性专业人员作出总结性的易用性测试报告。

行业通用格式（CIF）规范了用户测试过程中获取的信息类型。主要的可变因素是用户统计数据、任务描述、测试周境（包括所用设备、测试执行环境、受测人和测试管理员交互时使用的协议），以及为规范研究发现而选择的特别度量。

行业通用格式（CIF）目的是取代软件供应商和软件购买方进行易用性测试所使用的自有格式。使用标准报告格式的优点包括：(1)减少易用性测试人员的培训时间，因为不管一个人为多少家企业工作，他只需要学会使用一种格式；(2) 进一步提高供应商和购买方之间沟通效果，因为作为行业通用格式适用报告的读者，他们会有共同的语言和期望。

本文件旨在促使易用性成为交互性产品采购决策过程的一部分，以便更容易判断产品是否满足易用性目标。这种决策过程的例子包括购买、升级和采用自动化技术。它为供方的人类工效学工程师和易用性专业人员提供了一个向顾客描述易用性测试方法和结果的通用格式。

读者

供方组织中的易用性专业人员可利用行业通用格式（CIF）生成可供顾客组织使用的报告。它也供顾客组织确认一个特定报告是否符合行业通用格式。易用性测试报告本身针对以下两类读者：

- 顾客组织中评价易用性测试技术价值和产品易用性的专业人员；
- 其他专业人员和根据易用性测试结果进行商业决策的管理者。

单一组织也可使用行业通用格式来生成正式的易用性测试报告。在这种情况下，需要提供附加材料，如详细的发现清单。

报告包含执行摘要和正文两个主要部分。正文适用于上面提及的第一类读者，内容包括方法和结果。这些部分：(1)描述了测试方法和结果的技术细节，使得在测试可重复的情况下，允许其他组织重复测试过程；(2)支持使用测试数据权衡产品预期成本和收益。为了达到最佳效果，理解和解释这些部分需要人类工效学或易用性工程学技术背景。执行摘要和正文的引言部分为非易用性专业人员和管理者提供摘要信息，适用于第二类读者。当然，它也可能是其他计算机专业人员普遍关心的。没有易用性工程专业知识的决策者可在正文中找到有用信息，但必要时宜依靠专家解释。

软件工程 软件产品质量要求与 评价(SQuaRE) 易用性测试报告 行业通用格式(CIF)

1 范围

GB/T 25000 的本部分适用于按照 GB/T 18978.11—2004 对特定使用周境下的有效性、效率和满意度进行易用性测试后的报告编制工作。

注：易用性需求的其他度量可参考 GB/T 16260.2--2006、GB/T 16260.3—2006。

本部分适用于以下情况：

- 供方组织易用性专业人员编写供顾客组织使用的报告时；
- 顾客组织验证一个特定报告是否符合本文件时；
- 顾客组织内的人类工效学专家或其他易用性专业人员评价易用性测试的技术价值和产品易用性时；
- 顾客组织内的其他专业人员和管理者在利用测试结果对产品适宜性和购买进行商业决策时。

“5.2 执行摘要”和“5.3 引言”为非易用性专业人员和管理者提供了摘要信息。

5.4 和 5.5 从技术细节方面描述了测试方法和结果，也可通过这些测试数据权衡产品预期成本和收益。为了达到最佳效果，理解和解释这些章条，需要人类工效学或易用性工程的技术背景。

报告格式假设：在测试设计和执行过程中，遵循正规的做法^[1,2]。宜使用产生利于总结的易用性测试的测试规程，即此类测试是总结性的。某些易用性评价方法主要用于识别问题，而不在于产生测量值，如形成性测试。本部分不支持这种测试方法形成的结果。

2 符合性

如果易用性测试报告包含了本文件中要求的所有内容(描述为“应”)，则其符合本文件的要求。如果可能，推荐使用建议性要求(描述为“宜”)。

本部分规定了易用性测试报告宜提供的最基本信息。用户可根据需要进一步增加必要的信息。例如，如果某个组织发现其他的有用信息，那么即使行业通用格式中没有规定该部分内容，也可把这些信息加入报告中。

3 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 16260.1—2006 软件工程 产品质量 第1部分：质量模型(ISO/IEC 9126-1:2001, IDT)

GB/T 16260.2—2006 软件工程 产品质量 第2部分：外部度量(ISO/IEC TR 9126-2:2001, IDT)

GB/T 16260.3—2006 软件工程 产品质量 第3部分：内部度量(ISO/IEC TR 9126-3:2001, IDT)

GB/T 16260.4—2006 软件工程 产品质量 第4部分：使用质量的度量(ISO/IEC TR 9126-4:2001, IDT)

GB/T 18905.5—2002 软件工程 产品评价 第5部分:评价者用的过程(ISO/IEC 14598-5:1998,IDT)

GB/T 18976—2003 以人为中心的交互系统设计过程(ISO 13407:1999,IDT)

GB/T 18978.11—2004 使用视觉显示终端(VDTs)办公的人类工效学要求 第11部分:可用性指南(ISO 9241-11:1998,IDT)

4 术语和定义

GB/T 18978.11—2004 界定的以及下列术语和定义适用于本文件。

注:效率、有效性以及满意度在 GB/T 16260 中有定义,与 GB/T 18978 中的定义相似。不过在 GB/T 16260 中是从软件产品的角度去定义的,而在 GB/T 18978 中是以用户的观点去定义的。本文件中的术语“效率(efficiency)”等同于 GB/T 16260 中的术语“生产率(productivity)”,在 GB/T 16260 中,效率一词具有另外的含义。

4.1

易用性 usability

在特定的使用周境下,特定的用户,使用产品在有效性、效率和满意度方面达到规定目标的程度。

注:GB/T 16260 中定义的“使用质量”类似于 GB/T 18978 中定义的“易用性”。但是,由于“使用质量”可受任何质量特性影响,因此它比 GB/T 16260 中从易理解性、易学性、易操作性、吸引性和依从性几个方面定义的易用性的含义更加广泛。

4.2

有效性 effectiveness

用户实现特定目标的准确性和完备性。

4.3

效率 efficiency

与用户实现目标的准确性和完备性相应的资源消耗。

注:与软件“效率”的定义相比,易用性中“效率”的定义更接近于“生产率 productivity”的含义。

4.4

满意度 satisfaction

无不适感,并对产品使用持肯定态度。

4.5

使用周境 context of use

用户、任务、设备(硬件、软件和资料)以及使用某产品的物理和社会环境。

4.6

用户 user

与产品交互的个体。

4.7

用户组 user group

根据年龄、文化、专业技术等可能影响易用性的因素区分用户的子集。

4.8

目标 goal

用户与产品交互所产生的期望结果。

注:将与用户交互相关的特定目标称为“任务目标”可能更合适。

4.9

任务 task

实现目标所必需的活动。

注 1: 这些活动可能是体力的或认知上的。

注 2: 工作职责可决定目标和任务。

4.10

可访问性 accessibility

不管能力大小,人们使用产品、服务、环境或设施的易用性程度。

注 1: 尽管“可访问性 accessibility”一般是针对有缺陷的用户而言的,但这个概念并不限于在残疾范畴使用。

注 2: 面向易用性的可访问性概念主要关注在特定的使用周境下可获得的有效性、效率和满意度的级别,同时特别关注使用群体的能力差异,使个体的易用性体验的差异最小化。

4.11

援助技术 assistive technologies

为提高可访问性而添加到系统或与系统进行集成的硬件或软件。

示例:可发声的盲字输出设备、屏幕发声器、屏幕放大软件以及视线跟踪设备都是援助技术。

4.12

援助 assist

测试参与者无法自己完成任务时,测试管理者为了使测试继续而进行直接干预,为测试参与者提供帮助。

5 报告格式

5.1 标题页

应提供以下信息:

- 标明报告为 GB/T 25000.62 易用性测试报告行业通用格式,并提供联系信息;
- 命名被测产品和版本;
- 测试负责人;
- 测试执行时间;
- 报告起草日期;
- 报告起草人;
- 接受问题质疑和澄清时的联系人姓名;
- 供方联系电话;
- 供方 e-mail 地址;
- 供方邮寄地址。

宜提供以下信息:

- 顾客企业名称;
- 顾客企业联系人。

5.2 执行摘要

本条给出了测试的概述,旨在为顾客组织中的采购决策者提供信息,他们不必阅读文档的技术部分。

为便于本条作为一个独立的摘要使用,应在新页面开始,并以分页符结束。

应提供以下信息:

- 产品名称和描述;
- 方法摘要,包括参与者和任务的数量及类型;
- 以平均分或其他合适的集中趋势测度表示的执行结果和满意度。

宜提供以下信息：

- 测试的特性及目的；
- 执行结果的表格汇总；
- 当认为产品或数值之间存在差异时，差异非偶然发生的概率。

5.3 引言

本条提供了对产品和测试目标的描述。

5.3.1 产品描述

应提供以下信息：

- 正式的产品名称以及发行版本；
- 产品被评价的部分；
- 产品的意向用户群。

宜提供以下信息：

- 产品支持的援助技术；
- 预期的产品使用环境(包括物理环境和社会环境)的简短描述；
- 产品支持的用户工作类型。

5.3.2 测试目标

应提供以下信息：

- 包括所有指定兴趣领域的测试目标；

注：可能的目标包括，在使用产品时，测试用户的任务执行情况和主观满意度，旨在评估该产品是否符合指定的成功准则。

- 与用户直接或间接交互的功能和部件。

宜提供以下信息：

- 未对完整产品进行测试时，宜说明关注产品某个子集的原因。

5.4 方法

应给出详细信息，使得独立的测试人员可重复进行测试。

5.4.1 参与者

应提供以下信息：

- 测试的参与者总数；

注：为了进行有效的总结性的统计分析，应有足够数量的受测人，因此，推荐8个以上的受测人^[9]。

- 如果参与测试的用户组多于一个，应说明用户组的划分；
- 用户组的关键特征和能力；
- 选择参与者的依据；参与者是否具有基本的特征；
- 抽取的参与者与用户群之间的差异；

示例：实际用户可参加培训，而受测人是未经培训的。

- 包含人数统计、专业经验、计算机经验和援助技术需求的表格，其中行代表参与者，列代表特征。

注：需要援助技术的受测人可能是生理或认知方面有缺陷的个体。

示例：表1为参与者统计示例表，该表列出的是典型特征，并不一定包括测试参与者的每一种特征。

表 1 参与者统计示例表

•	性别	年龄	教育情况	职业/职责	专业经验	计算机经验	产品经验
参与者 1	•	•	•	•	•	•	•
参与者 2	•	•	•	•	•	•	•
参与者 n	•	•	•	•	•	•	•
注 1：在“性别”栏，应输入男性或女性。 注 2：在“年龄”栏，应输入参与者的实际年龄。如果不能提供准确年龄，则需说明成员的年龄范围（如 25～45 岁）或年龄段（如小于 18 岁，大于 45 岁）。 注 3：在“教育情况”栏，应输入所受全日制正规教育的年限（例如，一个高中毕业生有 12 年教育经历，而一个大学本科毕业生则有 16 年教育经历）。在某些情况下，教育程度可描述为“最高学历”。 注 4：在“职业/职责”栏，应输入用户使用该产品时的工作角色。应尽可能地提供角色的头衔。 注 5：在“专业经验”栏，应给出用户担任某角色的时间。 注 6：在“计算机经验”栏，应给出用户的相关背景，如用户使用该平台或操作系统的经历，和用户在产品领域的经历。这可能不止一列。 注 7：在“产品经验”栏，应给出之前曾接触过的本产品或同类产品的类型和持续时间。							

特征和能力的描述应足够完整，从而可招聘一组基本相似的参与者。

所选的特征和能力宜与产品易用性相关，应能使顾客确定参与者与用户群的相似程度。

宜提供以下信息：

- 需要提供援助技术的用户的描述。

参与者不宜来自于同一测试组织或供方组织。

5.4.2 产品测试使用周境

应提供以下信息：

- 评价时的使用周境与预期使用周境之间的任何重要差异。

5.4.2.1 任务

应提供以下信息：

- 测试的任务；
- 测试的任务场景；
- 选择任务的依据；

示例：最常使用的任务、最易发生错误的任务。

- 任务来源；

示例：使用类似产品的顾客的观察报告、产品市场需求规格说明和用户或设计团队之间的讨论报告。

- 提供给参与者的任务数据；

示例：数据入口的原始材料。

- 为每个任务确立的完成或执行的准则。

5.4.2.2 测试设施

宜提供以下信息：

- 进行评价的场地设置和类型；

示例：易用性实验室、办公室隔间、会议室、家庭办公室、寓所和生产车间。

- 任何可能影响结果的相关特征或环境。

示例：视频和音频记录设备、单面透视镜或自动数据收集设备。

5.4.2.3 参与者的计算环境

为进行重复测试和确认测试,应提供以下信息的详细情况,包括:

- 计算机配置,包括型号、操作系统版本、需要的程序库或设置;
- 如果使用了浏览器或相关插件,则说明其名称及版本。

5.4.2.3.1 显示设备

应提供以下信息:

- 如果产品用显示器输出,应说明屏幕的尺寸、分辨率和颜色设置;
- 如果产品用打印输出,应说明打印介质的尺寸和打印分辨率;
- 如果可视化界面元素(如字型)尺寸可变的话,应说明测试中所用的尺寸。

5.4.2.3.2 音频设备

宜提供以下信息:

- 如果产品有音频接口,宜说明音频比特(确定声音质量)、音量和其他属性的相关设置或值。

5.4.2.3.3 输入设备

宜提供以下信息:

- 如果使用了输入设备,宜说明测试用的设备的品牌和型号。

5.4.2.4 测试管理者工具

应提供以下信息:

- 如果使用了标准问卷(5.4.4.3 包含部分清单),应在此描述或详细说明。

宜提供以下信息:

用于控制测试或记录数据的任何硬件或软件。

5.4.3 实验设计

应提供以下信息:

- 测试的实验设计;为参与者分配实验条件的计划;以及与该计划的独立自变量和控制变量相关的统计分析。
- 独立自变量和控制变量;
- 为每一组条件记录数据所用的测度。

5.4.3.1 规程

应提供以下信息:

- 测度的操作定义;
- 独立自变量或控制变量的描述;
- 任务的时间限制;
- 测试人员与受测人间交互的策略和规程。

宜提供以下信息:

- 参与者从加入到解散的事件序列;
- 保密协议、完成形式、准备工作、预先培训和任务报告等细节;
- 验证参与者了解并理解他们作为受测人的权利^[1];
- 执行测试会话和记录数据的步骤;

- 测试会话期间与参与者进行沟通的人员数量和各自承担的角色；
- 测试环境中是否有其他个体在场；
- 是否向参与者支付酬金或提供其他补偿。

5.4.3.2 参与者一般性指导

应提供以下信息：

- 给予参与者的指导（在此描述或在附录中描述）；
- 参与者如何与在场的其他人进行沟通的指导，在适用的情况下，应包括他们如何请求援助和如何与其他参与者沟通。

5.4.3.3 参与者任务指导

应提供以下信息：

- 任务指导摘要。

5.4.4 易用性度量

如 4.1 中的定义，易用性是从有效性、效率和满意度这三种度量来进行测量的。测度选择应依赖于特定研究的目标、用户的特征、特定任务以及语境特征，示例参见参考文献[4]。

应提供以下信息：

- 有效性度量；
- 效率度量；
- 满意度度量。

即使在指定使用周境中难于解释，也应报告有效性和效率的结果。这时，报告应详述供方认为这些度量没有意义的原因。

示例：假定，产品的使用周境中包括在密切相关事务间的实时的、开放式交互。在这种情况下“任务占用时间”解释为效率测度或许没有意义。因为对许多用户来说，花费在该任务上的时间是“良好的开销”。

如果必须为参与者提供援助，那么应同时提供受援时和非受援时的有效性和效率度量，所提供援助的数量和类型也应作为测试结果的一部分包括其中。

5.4.4.1 有效性

有效性与使用该产品的目标和这些目标可达到的准确度及完备性有关。一般的有效性测度包括任务完成的百分比、出错频率、测试人员援助参与者的频率和在执行任务期间参与者访问文档或寻求帮助的频率。它只考虑任务目标的完成程度，而不考虑如何达到任务目标。效率使达到的有效性等级与开销的资源数量有关。

5.4.4.1.1 完成率

完成率是指完整且正确地完成了每一项任务目标的参与者比例。如果目标能部分完成（如未完成或未达到最佳结果），则通过与部分结果值相关的指定准则按 0~100% 记分，由此来报告所达到的平均目标。如拼写校对任务涉及识别和校正十个拼写错误，则完成率可根据纠错的百分率计算。另一个计算完成率的方法是权值法，如在文档标题页中的拼写错误的重要程度是文本主体中的两倍。如果在报告中包括这种部分完成的结果，则宜说明进行部分目标分析时所选择的特定方法的基本原理。

应提供以下信息：

- 完整且正确地完成了每一项任务目标的参与者比例。

5.4.4.1.2 差错

差错是指测试参与者未成功完成任务，或不得不多次尝试执行部分任务。数据评分宜包括按照某

种分类法对差错进行分类,如参考文献[10]。

5.4.4.1.3 援助

当参与者不能继续一个任务时,有时测试管理者会直接给予程序上的帮助,以使测试继续进行。尽管“援助”测度与易学性和有效性均有关联,但为了与 GB/T 18978.11—2004 中定义的易用性的三个元素相符合,本部分中将其与有效性结合。

应提供以下信息:

- 非受援完成率(即在测试人员未干涉的情况下的完成率);
- 在受援完成率和非受援完成率有差别的地方,受援完成率(即在测试人员干涉情况下的完成率)。

示例:如果参与者在任务 A 中接受了援助,在计算该任务的非受援完成率时,则该参与者不宜包括在成功完成该任务的参与者中。但如果接受援助后成功完成了该任务,则在计算任务 A 的受援完成率时,可把他包括在内。

当允许或提供援助时,应在测试结果报告中说明援助的类型和次数。如果研究设计允许援助,那么提供受援完成率和提供非受援完成率没有太大的差别。在这种情况下,只需要提供一个完成率,但应包括援助的类型和次数。

示例:在教育方面的设置中,鼓励学生来寻求援助。

在一些易用性测试中,当参与者不能独立完成任务时,他们可得到帮助工具(如在线帮助或文档)的指导,而这些工具是产品的一部分。对本报告来说,获取提示信息 and 在线帮助的产品特性不算是一种援助。不过,如果产品的帮助特性可使参与者获得独立使用产品的能力,那么还是应在报告中记录获取不同产品帮助特性的频率。

5.4.4.2 效率

效率与达到的有效性级别以及开销的资源数量有关。效率通常是用完成任务的平均时间来评估的。效率也与其他资源相关(如总开销)。效率的一般测度是完成任务的时间。

应提供以下信息:

- 参与者完成每项任务的平均时间,以及时间范围和标准差。

5.4.4.2.1 完成率/任务时间均值

完成率/任务时间均值是效率的另一个测度^[10]。通过成功率与时间的关系,顾客可将快速但易错的界面(如用通配符批量删除文件的命令行界面)与慢速但易用的界面(如用鼠标和键盘把文件逐个拖至回收站的界面)进行对比。

5.4.4.3 满意度

应提供以下信息:

- 一个或多个用户满意度的测度。

满意度是指使用产品时,用户的主观反应。用户满意度可能和使用产品的动机相关,并且在某些情况下可能影响使用性能。用于测量满意度和相关看法的问卷,通常使用 Likert 标度和语义差异标度来构造。

可使用不同的工具测量交互式软件产品的用户满意度,也有许多组织创建了自己的工具。不管是使用外部的、标准的工具,还是自己定制的工具,宜包含主观评级方法,如满意度、有用性和易使用性,因为顾客组织普遍对此感兴趣。

许多问卷被广泛使用,包括 ASQ^[6]、CSUI^[7]、PSSUQ^[6]、QUIS^[3]、SUMI^[5] 和 SUS^[2]。尽管每种问卷都给出了一些对产品易用性主观测度的独特看法,但大多数都包括满意度、有用性和易使用性的测量。参考文献[4]对问卷结构的一般信息进行了解释。

供方可选择使用已发布的、经确认的满意度测度或提交他们自己完成的满意度度量。

5.5 结果

5.5.1 数据分析

应对以下信息做详细的描述,以便于在重复测试时,其他组织可重复所使用的数据分析方法:

- 数据采集:规划要采集的数据和实际采集的数据之间的差异。
示例:如何处理缺失数据;排除局外人的情况下,如何处理数据。
- 数据评分:收集到的数据值和后续分析中使用的值之间的对应关系;
示例:差错如何分类;实际年龄和年龄段的对应关系;受援差错和一系列值之间的对应关系。
- 数据简化:计算集中趋势测度和描述数据变化的方法;
示例:使用哪种集中趋势测度(如平均值或模式);如何测量变化(如标准差或范围)。
- 统计分析:分析数据的统计规程。
示例:组之间如何进行比较(如 t-测试, F-测试等)。

所报告的平均值数据应包括标准差和可选的均值的标准误差。

5.5.2 结果表述

应提供以下信息:

- 用表格形式表示的每个任务或任务组执行结果的有效性和效率的测度。

示例:可仿效表 2 来描述任务组中的每个任务。

表 2 任务 A

用户 u	非受援情况下的任务 有效性 [(%)完成量]	受援情况下的任务 有效性[(%)完成量]	任务时间 (分)		差错	援助	效率
1							
2							
N							
平均值							
标准差							
最小值							
最大值							

不同的图形格式有助于表述易用性数据。条形图用来表示主观数据,如利用 Likert 标度方法采集的数据。通过有效使用不同的图形能对一个产品的专家基准时间和参与者执行的平均时间进行对比。这些数据可附有结果的简短说明,但不鼓励在报告的正文中对此做详细解释。在必要时,可在附录中进行详细解释。

可提供以下信息:

- 描述执行结果的有效性和效率测度的图。

5.5.2.1 执行结果

宜提供以下信息:

- 描述所有任务执行结果的有效性和效率测度的汇总表;
- 描述执行结果的有效性和效率测度的图形表示;
- 如果度量和产品设计及特定应用领域相关,应提供附加的度量表。

示例:表 3——所有任务执行结果的汇总表。

表 3 任务执行结果汇总表

用户 #	非受援情况下的任务 有效性总量 [(%)完成量]	受援情况下的任务 有效性总量[(%)完成量]	总任务时间 (分)		差错总数	援助总数	效率
1							
2							
N							
平均值							
标准差							
最小值							
最大值							

结果表格可用来表示相关任务组(例如,一个组包括所有的程序编制任务,另一个组包括所有的调试任务),这会更有效、更有意义。如果单元任务有子任务,那么可以以摘要的方式记录子任务。

示例:如果单元任务的目的是识别一页中所有写错的字,那么结果可概括为已发现错误书写的百分比。

5.5.2.2 满意度结果

应提供以下信息:

——描述满意度结果的测度的汇总表。

示例:表 4——部分满意度结果的汇总表。

表 4 满意度汇总表

参与者编号	易使用性	有用性	外观	清晰度	
1					
2					
N					
中间值					
最小值					
最大值					

宜提供以下信息:

——描述满意度结果的测度的图形表示。

5.6 附件

应提供以下信息:

——所使用的定制问卷(见 5.4.2.4 和 5.4.4.3);

——参与者一般性指导(若报告正文中未提供)(见 5.4.3.2);

——参与者任务指导(见 5.4.3.3)。

宜提供以下信息:

——发布说明或升级测试结果的注释。

附 录 A
(资料性附录)
检 查 表

本附录给出了易用性测试报告的检查表,见表 A.1。

表 A.1 检查表

使用以下清单确保您的行业通用格式报告中包含了必需元素(用※表示)。推荐项用◆表示。	
标题页	
◆公司标志或名称	
※标明报告为 GB/T 25000.62 易用性测试行业通用格式	
※命名被测产品和版本	
※测试负责人	
※测试执行时间	
※报告起草日期	
※报告起草人	
◆顾客企业名称	
◆顾客企业联系人	
※接受问题质疑和(或)澄清时的联系人姓名	
※供方联系电话	
※供方 e-mail 地址	
※供方邮寄地址	
执行摘要	
※提供测试的总体概述	
※产品名称和描述	
※方法摘要,包括参与者和任务的数量及类型	
※以平均分或其他合适的集中趋势测度表示的结果	
◆测试的特性及目的	
◆执行结果的表格汇总	
◆如果声明了差异,应说明相关统计概率	
※在新的一页中开始;以分页符结束	
引言	
产品描述	
※正式的产品名称以及发行版本	
※描述产品的被评价部分	
※产品的意向用户群	
◆产品支持的援助技术	
◆宜使用产品的环境的简述	
◆产品支持的用户工作类型	

表 A.1 (续)

测试目标
※陈述测试目标及任何指定兴趣领域
※与用户直接或间接交互的功能和部件
◆关注产品某个子集的原因
方法
参与者
※测试的参与者总数
※如果测试的用户组多于一个,应说明用户组的划分
※用户组的关键特征和资质
※选择参与者的方法;参与者是否具有基本的特征
※抽取的参与者与用户群之间的差异
※参与者(行)特征(列)的表格
◆需要提供援助技术的用户的描述
产品测试使用周境
※评价的使用周境与预期使用周境之间的任何重要差异
任务
※描述被测试的任务
※描述测试任务的场景
※解释选择任务的依据
※描述任务来源
※提供给参与者的任务数据
※为每个任务确立的完成准则或性能准则
测试设施
◆描述进行评价所用场地的设置和类型
◆详述任何可能影响结果的相关特征或周围环境
参与者的计算环境
※计算机配置,包括型号、操作系统版本、所需要的程序库或设置
※如果使用了浏览器或相关插件,应说明其名称及版本
显示设备
※如果产品用显示器输出,应说明屏幕的尺寸、分辨率和颜色设置
※如果产品用打印机输出,应说明打印介质尺寸和打印分辨率
※如果可视化界面元素(如字体)尺寸可变的话,应说明测试中所用的尺寸
音频设备
◆如果产品有音频接口,宜说明音频比特、音量和其他属性的相关设置或属性值
手工输入设备
◆如果使用了输入设备,宜说明测试用的设备的品牌和型号
测试管理者工具
※如果使用了标准问卷,在此描述或详细说明
◆描述用于控制测试或记录数据的任何硬件或软件
实验设计
※描述对测试实验的设计
※定义独立自变量和控制变量

表 A.1 (续)

※描述记录数据所用的测度
规程
※测度的操作定义
※独立自变量和控制变量的描述
※任务的时间限制
※测试人员与受测人间沟通的策略和规程
◆参与者从加入到解散的事件序列
◆保密协议、完成表格的填写、准备工作、预先培训和任务报告
◆确认参与者了解并理解他们作为受测人的权利
◆说明执行测试和记录数据的步骤
◆测试期间与参与者进行沟通的人员数量和各自承担的角色
◆说明测试环境中是否有其他个体在场
◆陈述是否向参与者支付酬金
参与者一般性指导
※给予参与者的指导(在此描述或在附录中描述)
※参与者如何与在场的其他人进行沟通的指导
参与者任务指导
※任务指导摘要
易用性度量
※有效性度量
※效率度量
※满意度度量
结果
数据分析
※数据采集
※数据评分
※数据简化
※统计分析
结果表述
※用表格形式表示的每个任务或任务组执行结果的有效性和效率的测度
执行结果
◆所有任务执行结果的汇总表
◆执行结果的图形表示
◆附加的度量表(若相关的话)
满意度结果
※描述满意度结果的测度汇总表
※描述满意度结果的测度的图形化表示
附件
※所使用的定制问卷
※参与者一般性指导
※参与者任务指导
◆对发布的注释

附 录 B
(资料性附录)
术 语

B.1

自动数据收集设备 automatic data collection equipment

自动采集易用性测试数据的硬件和/或软件。数据采集的例子包括屏幕捕获,视频、音频和键盘记录,而且通常伴有时间戳。

B.2

配置 configuration

组成产品测试计算环境的指定硬件和软件,包括型号和版本号。配置包括平台。

B.3

顾客 customer

打算购买产品的组织。

B.4

任务报告 debriefing

测试后对参与者的访谈。

B.5

实验设计 experimental design

将实验条件分配给参与者的计划,以及与此计划关联的统计分析。

通常他包括对独立自变量、因变量、参与者人数和取样策略的说明,将实验条件分配给参与者的规程,以及指派测试任务的顺序。

B.6

形成性评价 formative evaluation

一种测试方法,它是迭代式设计过程的一部分,通常涉及“边想边说规程”的使用(与总结性测试相对)。

B.7

独立自变量 independent variable

测试人员用来建立测试条件所使用的变量(与因变量相对,它是研究中需要测量的变量)。

示例:如果对两个系统中的相同任务进行比较,则独立自变量为系统,因变量为任务执行水平或用户满意程度。

B.8

Likert 标度 Likert scale

通过量化主观信息来测量用户态度或反应的评定标度。它通常包含一组关于产品的褒贬不一的陈述,受访者需表明他(她)对这些陈述赞同或不赞同的程度。一个 Likert 问卷可有奇数或偶数个要点,通常用数字进行编号,并且通常至少有两个锚来描述标度的极端情况。

示例:

该产品易于使用

强烈不同意	基本不同意	稍微不同意	中立	稍微同意	基本同意	强烈同意
-------	-------	-------	----	------	------	------

使用该产品时我感觉很沮丧

强烈不同意	基本不同意	稍微不同意	中立	稍微同意	基本同意	强烈同意
-------	-------	-------	----	------	------	------

B.9

平均数 mean

一组值的算术平均。具体来说,即以一组值的和除以这组值的数量。

B.10

保密协议 nondisclosure agreement

供方与顾客组织签订的协议,协议中声明他们的合作需要将某些专有资料披露给对方,并同意不将对方的专有资料泄露给其他任何人。

B.11

单向镜 one-way mirror

一种玻璃,依据照明条件的不同,一面呈现为镜子而另一面呈现为窗口。该设备可让测试人员在自己不被看到或不影响参与者的情况下观察易用性测试。

B.12

参与者 participant

作为产品目标用户群的代表参与易用性测试的人。参与者执行测试任务并遵循测试人员给出的其他指导。

B.13

平台 platform

用于测试的某类计算机(操作系统和硬件)。常见平台包括基于 IdeaCentre 的 Windows,基于 iMac 的 MacOS,基于 Sparc 的 UNIX,基于 Dell 的 Linux,以及基于 IBM 的 OS/2。时尚的设备如掌上电脑也可视为平台。(注:此处的平台信息根据市场和国情做了修改)

B.14

群 population

从中选取统计样本的一组个体。

B.15

采购决策者 procurement decision maker

负责决策顾客组织购买哪些产品的人。

B.16

重复 replication

在相同条件下对两次观察情况的收集。本文件中,它是指顾客组织宜能够利用自己的参与者来重复报告中描述的测试。

B.17

场景 scenario

以用户角度描述的,在指定周境下执行任务的事件序列。

B.18

语义差异 semantic differential

通过量化主观信息来测量用户态度或反应的评定量表。它通常包含一组描述产品某些特性的两极性形容词组(例如,好/坏,易/难)。语义差异问卷通常设有 7 到 10 级。

示例:

该产品

易于使用	1	2	3	4	5	6	7	难以使用
令人沮丧的	1	2	3	4	5	6	7	令人满意的

B.19

标准差 standard deviation

方差的平方根。它为衡量测量分布的可变性提供了一种测度。

B.20

平均数标准误差 standard error of the mean

(方差除以样本大小)的平方根。它可估计样本平均值和总平均值之间的近似性。

B.21

具有统计学意义的 statistically significant

非偶然发生的(即,有可能反映实际差异)。

B.22

总结性测试 summative test

对完整产品进行的易用性测试,以确定它满足易用性目标的程度。

B.23

供方 supplier

产品的生产者。为顾客提供产品的组织(ISO/IEC TR 15504-9:1998)

B.24

任务场景 task scenario

实际可行、简短的任务介绍和描述。

B.25

分类 taxonomy

在某一研究领域中对研究对象的系统分类。

B.26

测试协议 test protocol

测试中应遵循的步骤列表。

B.27

测试人员/测试管理员 tester/test administrator

设计和管理易用性测试的人(测试人员不是与产品进行交互,执行任务场景的人;这样的人被称为测试参与者)。

B.28

易用性专业人员/专家/工程师 usability professional/specialist/engineer

从事易用性工程/测试实践的人。这些从业者懂得如何设计和执行以人为测试对象的总结性性能测试,以及如何分析结果数据。他们的经验通常在人类工效学、行为科学、工业工程、人机交互、工业设计、计算机科学或相关领域处于领先地位。

B.29

确认 validation

对诸如问卷或易用性测试等测度的有效性的确定。

B.30

方差 variance

一组测量和它们的平均值之差的平方和,除以样本大小(参见标准差)。

附录 C
(资料性附录)
报告模板

GB/T 25000.62 易用性测试报告行业通用格式

[※¹⁾命名被测产品和版本]

[※ 测试负责人]

[※ 测试执行时间]

[※ 报告起草日期]

[※ 报告起草人]

[◆ 顾客企业名称]

[◆ 顾客企业联系人]

问题致信:[※ 接受问题质疑和澄清时的联系人姓名]

电 话:[※ 供方联系电话]

电子邮箱:[※ 供方 e-mail 地址]

联系地址:[※ 供方邮寄地址]

1) 图形符号说明:※表示必填信息项,◆表示推荐填写信息项。

执行摘要

- [※ 提供测试的总体概述]
- [※ 命名和描述产品]
- [※ 方法摘要,包括参与者和任务的数量及类型]
- [※ 以平均分或其他合适的集中趋势测度表示的结果]
- [◆ 测试的特性及目的]
- [◆ 性能测试结果汇总表]
- [◆ 如果声明了差异,应说明相关统计概率]

引言

产品描述

- [※ 正式的产品名称以及发行版本]
- [※ 描述产品被评价的部分]
- [※ 产品的意向用户群体]
- [◆ 产品支持的援助技术]
- [◆ 产品预期使用环境的简短描述]
- [◆ 产品支持的用户工作类型]

测试目标

- [※ 陈述测试目标及任何指定兴趣领域]
- [※ 用户直接或间接与之交互的功能和部件]
- [◆ 关注产品某个子集的原因]

方法

参与者

- [※ 测试参与者的总数]
- [※ 如果有多个参与测试的用户组,应说明用户组的划分]
- [※ 用户组的关键特征和资质]
- [※ 选择参与者的方法;参与者是否具有基本的特征]
- [※ 抽取的参与者与用户群体之间的差异]
- [※ 参与者(行)与特征(列)的表格]

	特征 1	特征 2		特征 N
参与者 1				
参与者 2				
.....				
参与者 N				

- [◆ 对需要援助技术的用户的描述]

产品测试使用周境

[※ 被评价的使用周境与预期的使用周境之间的任何已知差异]

任务

[※ 描述测试的任务]

[※ 描述测试的任务场景]

[※ 解释选择任务的依据]

[※ 描述任务来源]

[※ 提供给参与者的任务数据]

[※ 为每个任务确立的完成准则或性能准则]

测试设施

[◆ 描述进行评价所用场地的设置和类型]

[◆ 详述任何可能影响结果的相关特征或周围环境]

参与者的计算环境

[※ 计算机配置,包括型号、操作系统版本、所需要的程序库或设置]

[※ 如果使用了浏览器或相关插件,应说明其名称及版本]

显示设备

[※ 如果产品使用显示器输出,应说明屏幕的尺寸、分辨率和颜色设置]

[※ 如果产品使用打印机输出,应说明打印介质尺寸和打印分辨率]

[※ 如果可视化界面元素尺寸可变的话,应说明测试中所用的尺寸]

音频设备

[◆ 如果产品有音频接口,宜说明音频比特、音量和其他属性的相关设置或属性值]

手工输入设备

[◆ 如果使用了输入设备,宜说明测试用的设备的品牌和型号]

测试管理者工具

[※ 如果使用了标准问卷,在此描述或详细说明]

[◆ 描述用于控制测试或记录数据的任何硬件或软件]

实验设计

[※ 描述对测试实验的设计]

[※ 定义独立自变量和控制变量]

[※ 描述记录数据所用的测度]

规程

[※ 测度的操作定义]

[※ 独立自变量和控制变量的描述]

[※ 任务的时间限制]

[※ 测试人员与受测人间沟通的策略和规程]

[◆ 参与者从被组织到被解散的事件序列]

- [◆ 保密协议、完成表格的填写、准备工作、预先培训和任务报告]
- [◆ 确认参与者了解并理解他们作为受测人的权利]
- [◆ 说明执行测试和记录数据的步骤]
- [◆ 测试期间与参与者进行沟通的人员数量和各自承担的角色]
- [◆ 说明测试环境中是否有其他个体在场]
- [◆ 陈述是否向参与者支付酬金]

参与者一般性指导

- [※ 给予参与者的指导]
- [※ 关于参与者如何与在场的其他人进行沟通的指导]

参与者任务指导

- [※ 任务指导摘要]

易用性度量

- [※ 有效性度量]
- [※ 效率度量]
- [※ 满意度度量]

结果

数据分析

- [※ 数据采集]
- [※ 数据评分]
- [※ 数据简化]
- [※ 统计分析]

结果表述

- [※ 用表格形式表示的每个任务或任务组执行结果的有效性和效率的测度]

用户	测度 1	测度 2		测度 N
参与者 1				
参与者 2				
.....				
参与者 N				
平均值				
标准差				
最大值				
最小值				

执行结果

[◆ 描述所有任务执行结果的有效性和效率测度的汇总表]

用户	总测度 1	总测度 2		总测度 N
参与者 1				
参与者 2				
.....				
参与者 N				
平均值				
标准差				
最大值				
最小值				

[◆ 执行结果的图形表示]

[◆ 附加的度量表]

满意度结果

[※ 描述满意度结果的测度汇总表]

用户	标度 1	标度 2		标度 N
参与者 1				
参与者 2				
.....				
参与者 N				
平均值				
标准差				
最大值				
最小值				

[◆ 满意度结果的图形表示]

附件

[※ 所使用的定制问卷]

[※ 参与者一般性指导]

[※ 参与者任务指导]

[◆ 发布说明]

参 考 文 献

- [1] American Psychological Association. Ethical Principles in the Conduct of Research with Human Participants, 1982.
- [2] Brooke, J. (1996). SUS: A "quick and dirty" usability scale, In Jordan, P. W., Thomas, B. T., and Weerdmeester, B. A. (Eds.), Usability Evaluation in Industry. UK: Taylor and Francis, pp. 189-194.
- [3] Chin, J. P., Diehl, V. A., and Norman, K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In: CHI'88. Conference Proceedings on Human Factors in Computing Systems. New York: ACM, pp. 213-218.
- [4] Dumas, J., and Redish, G. (1999). A Practical Guide to Usability Testing. UK: Intellect.
- [5] Kirakowski, J. (1996). The software usability measurement inventory: Background and usage. In Jordan, P., Thomas, B., and Weerdmeester, B. (Eds.), Usability Evaluation in Industry. UK: Taylor and Francis, pp. 169-177.
- [6] Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies; the ASQ. SIGCHI Bulletin, 23(1), 78-81.
- [7] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7, 57-78.
- [8] Macleod, M., Bowden, R., Bevan, N. and Curson, I. (1997). The MUSiC Performance Measurement method. Behaviour and Information Technology, 16, 279-293.
- [9] Nielsen, J., and Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In: CHI'93. Conference Proceedings on Human Factors in Computing Systems. New York: ACM, pp. 206-213.
- [10] Norman, D. A. (1983). Design rules based on analyses of human error. Communications of the ACM, 26(4), 254-258.
- [11] Rubin, J. (1994). Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. New York: John Wiley and Sons, Inc.
-

中 华 人 民 共 和 国
国 家 标 准
软件工程 软件产品质量要求与
评价(SQuaRE) 易用性测试报告
行业通用格式(CIF)

GB/T 25000.62—2014/ISO/IEC 25062:2006

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100029)
北京市西城区三里河北街16号(100045)

网址 www.spc.net.cn

总编室:(010)64275323 发行中心:(010)51780235

读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 1.75 字数 48 千字
2014 年 9 月第一版 2014 年 9 月第一次印刷

*

书号: 155066·1-49357 定价 27.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



GB/T 25000.62-2014