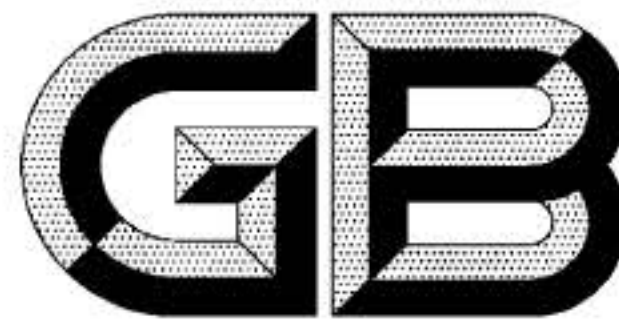




中华人民共和国国家标准

GB/T 25000.22—2019
代替 GB/T 16260.4—2006

系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第22部分:使用质量测量

Systems and software engineering—
Systems and software Quality Requirements and Evaluation (SQuaRE)—
Part 22: Measurement of quality in use

(ISO/IEC 25022:2016, Systems and software engineering—
Systems and software quality requirements and evaluation (SQuaRE)—
Measurement of quality in use, MOD)

2019-08-30 发布

2020-03-01 实施

国家市场监督管理总局 发布
中国国家标准化管理委员会

目 次

前言 I

引言 III

1 范围 1

2 符合性 1

3 规范性引用文件 2

4 术语和定义 2

5 缩略语 5

6 使用质量测度的使用 5

7 质量测度记录格式 8

8 使用质量测度 8

附录 A（资料性附录） 本部分与 ISO/IEC 25022:2016 相比的结构变化情况 20

附录 B（资料性附录） 本部分与 ISO/IEC 25022:2016 的技术性差异及其原因 21

附录 C（资料性附录） 周境覆盖测量方法示例 22

附录 D（资料性附录） 使用质量测度规范化 24

附录 E（资料性附录） 使用本部分测量 GB/T 18978.11—2004 中的易用性 28

附录 F（资料性附录） 使用质量评价过程 29

附录 G（资料性附录） 不同质量模型之间的关系 34

附录 H（资料性附录） 质量测量概念 35

附录 I（资料性附录） 用于定义质量测度的质量测度元素 36

参考文献 37

前 言

GB/T 25000《系统与软件工程 系统与软件质量要求和评价(SQure)》分为以下部分:

- 第 1 部分:SQure 指南;
- 第 2 部分:计划与管理;
- 第 10 部分:系统与软件质量模型;
- 第 12 部分:数据质量模型;
- 第 20 部分:测量参考模型和指南;
- 第 21 部分:质量测度元素;
- 第 22 部分:使用质量测量;
- 第 23 部分:系统与软件产品质量测量;
- 第 24 部分:数据质量测量;
- 第 30 部分:质量需求;
- 第 40 部分:评价过程;
- 第 41 部分:开发方、需方和独立评价方的评价指南;
- 第 45 部分:可恢复性的评价模块;
- 第 51 部分:就绪可用软件产品(RUSP)的质量需求和测试细则;
- 第 62 部分:易用性测试报告行业通用格式(CIF)。

本部分为 GB/T 25000 的第 22 部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

本部分代替 GB/T 16260.4—2006《软件工程 产品质量 第 4 部分:使用质量的度量》,与 GB/T 16260.4—2006 相比,主要技术变化如下:

- 依据 GB/T 25000.10 中修订的使用质量模型,将具有 4 个特性且无子特性的使用质量的度量修改为具有 5 个特性及若干子特性的使用质量测度,修订细节如下:
 - 修改有效性度量的名称为有效性测度(见 8.2,2006 年版的 8.1);
 - 修改生产率度量的名称为效率测度(见 8.3,2006 年版的 8.2);
 - 删除了安全性度量(见 2006 年版的 8.3);
 - 增加了抗风险测度及其 3 个子特性的测度(经济风险缓解性测度、健康和安全风险缓解性测度、环境风险缓解性测度)(见 8.5);
 - 修改满意度度量的名称为满意度测度,并增加了其 4 个子特性的测度(有用性测度、可信性测度、愉悦性测度、舒适性测度)(见 8.4,2006 年版的 8.4);
 - 增加了周境覆盖测度及其 2 个子特性的测度(周境完备性测度和灵活性测度)(见 8.6);
- 将测度归类为通用测度(G)和特定测度(S)(见第 7 章);
- 给出的使用质量测度表包括 ID、名称、描述、测量函数和方法,修改简化了 GB/T 16260.4—2006 度量表的内容(见第 8 章表 1~表 13,2006 年版的表 8.1~表 8.4)。

本部分使用重新起草法修改采用 ISO/IEC 25022:2016《系统与软件工程 系统与软件质量要求和评价(SQure) 使用质量测量》。

本部分与 ISO/IEC 25022:2016 相比在结构上有少量调整,附录 A 中列出了本部分与 ISO/IEC 25022:2016 的章条编号对照一览表。

本部分与 ISO/IEC 25022:2016 相比存在技术性差异,附录 B 中给出了相应技术性差异及其原因

的一览表。

本部分还做了下列编辑性修改：

- 为与 GB/T 25000 系列国家标准相协调,标准名称中增加了“第 22 部分”；
- 修改了 4.13 定义中的“质量测量元素”(quality measurement element)为“质量测度元素”(quality measure element),使其与 4.17 的术语名称保持一致；
- 在附录 D 中,修改测度 SUs-1-G、SUs-2-G、RHe-2-G 和 CF1-2-S 的测量函数公式的写法,使其分别与正文第 8 章表 3、表 4、表 10 和表 13 中的相同公式在写法上保持一致；
- 在附录 D 中,修改测度 CF1-2-S 的测量函数中的“每个新需求”为“第 i 个需求”,使其与正文 8.6.3 表 13 中的相同内容在表述上保持一致；
- 删除了与规范性引用文件重复的参考文献；
- 参考文献增加了被 8.6.3 注 2 参考的“ISO 9241-110:2006 Ergonomics of human-system interaction—Part 110: Dialogue principles”。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

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、中国电子技术标准化研究院、中国司法大数据研究院有限公司、国家应用软件产品质量监督检验中心、中国合格评定国家认可中心、上海计算机软件技术开发中心、北京梆梆安全科技有限公司、北京东方国信科技股份有限公司、南京大学、中电莱斯信息系统有限公司、厦门理工学院、中国电子科技集团公司第五十四研究所、山东道普测评技术有限公司。

本部分主要起草人:张旸旸、邓姿娴、王珩、黄家裕、刘潇健、郭新伟、郭澍、吴克寿、李彦军、蔡立志、秦晓秋、韦晓枝、庞懿丽、卢梁春、胡晓冉、赵毅、秦泗强、韩庆良、赵红卫、赵宏博、卢俊文、崔建峰、陈振宇、房春荣、周晓明、孙纪敏、郭晓珍、康京山、李浩文、韩明军、刘振宇、龚家瑜、胡芸。

本部分所代替标准的历次版本发布情况为：

- GB/T 16260.4—2006。

引 言

GB/T 25000 的本部分提供了一组针对(GB/T 25000.10 中定义的)使用质量特性的测度,这些测度能够(结合 GB/T 25000.30)用于规定使用质量需求,以及(结合 GB/T 25000.40 和 GB/T 25000.41)用于测量和评价使用质量。

本部分基于实用价值来选取质量测度。这些测度是以现有实践(例如参考文献[10])为基础的。这些测度并不是全面详尽的,鼓励本部分的用户在必要时加以完善。

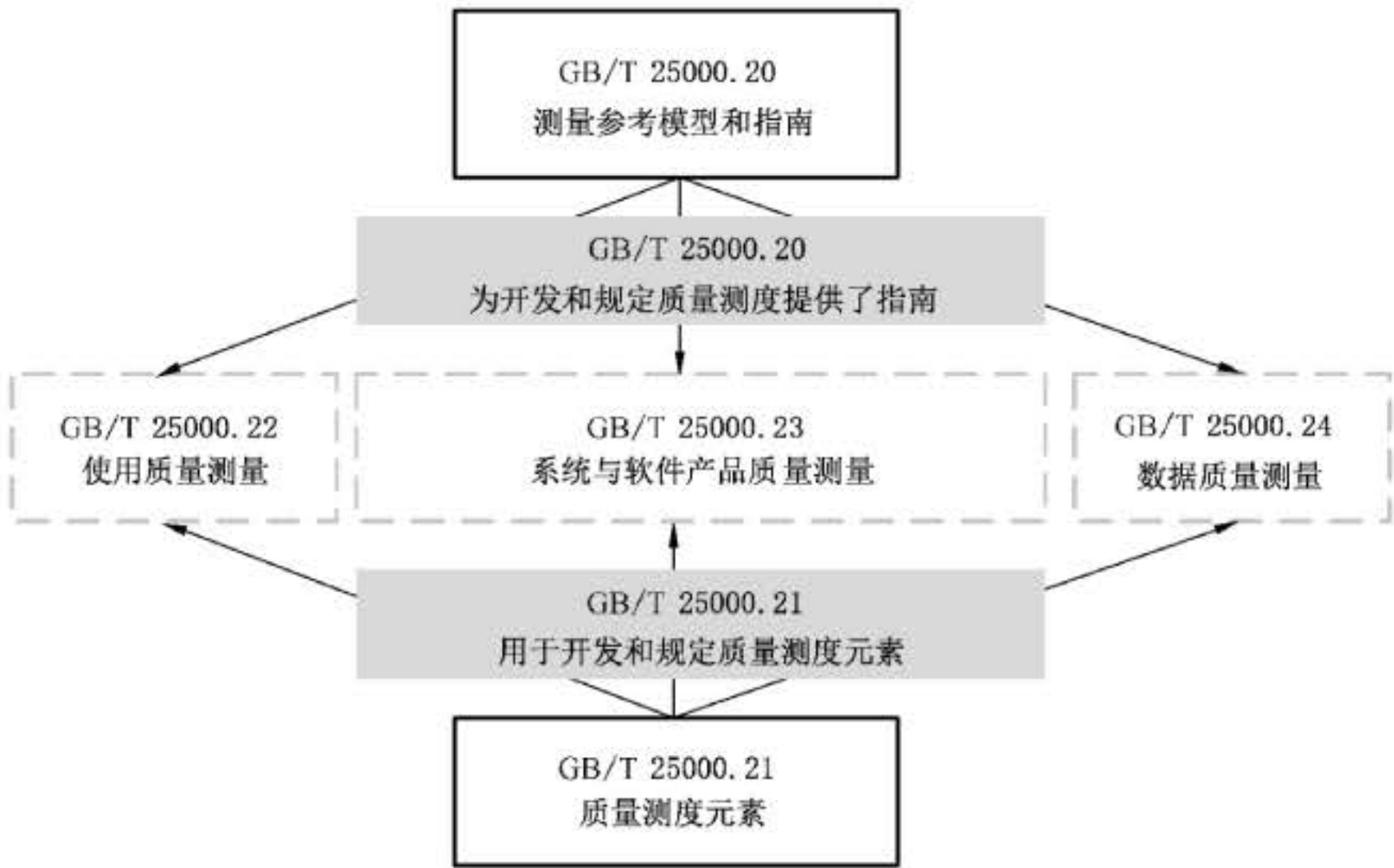


图 1 SQuaRE 质量测量标准的关系

图 1 描绘了 GB/T 25000.22 与其他 SQuaRE 质量测量标准的关系：

- a) GB/T 25000.20——测量参考模型和指南：提供用于测量质量特性的参考模型和指南；
- b) GB/T 25000.21——质量测度元素：提供用于规定质量测度元素的一种格式,以及能够用于构建软件质量测度的质量测度元素的若干示例；
- c) GB/T 25000.22——使用质量测量：提供测度,包括使用质量模型中的质量特性的关联测量函数；
- d) GB/T 25000.23——系统与软件产品质量测量：提供测度,包括产品质量模型中的质量特性的关联测量函数和质量测度元素；
- e) GB/T 25000.24——数据质量测量：提供测度,包括数据质量模型中的质量特性的关联测量函数和质量测度元素。

系统与软件工程

系统与软件质量要求和评价(SQuaRE)

第 22 部分:使用质量测量

1 范围

GB/T 25000 的本部分针对 GB/T 25000.10—2016 所定义之特性的使用质量测度进行了定义,旨在与 GB/T 25000.10—2016 搭配使用。本部分能与 GB/T 25000.30、GB/T 25000.40 和 GB/T 25000.41 等标准结合使用,并能在产品或系统质量方面更普遍地满足用户需要。

本部分包含如下内容:

- 针对每个使用质量特性的基本测度集;
- 关于如何测量使用质量的解释说明。

本部分提供了一组与 GB/T 25000.10—2016 中的使用质量模型一起使用的推荐性的使用质量测度,这些测度并不是全面详尽的。

这些测度可应用在任何人机系统的使用中,既包括使用中的计算机系统,也包括作为系统一部分的软件产品。

本部分不指定达到评定等级或依从性等级的测度值域,因为每个系统或产品的值取决于使用周境和用户需要而定。

一些属性可能有理想值范围,这些值取决于例如人类认知因素的一般因素,而不是特定用户需要。

本部分提出的使用质量测度主要在基于真实使用效果的系统与软件产品的质量保证和管理中使用。测量结果的主要用户是软件与系统开发、获取、评价或维护的管理人员。

本部分的主要用户是开展规约和评价活动的人员,这些活动包括:

- 开发:包括生存周期过程中的需求分析、设计和验收测试;
- 质量管理:对产品或计算机系统的系统性检查,例如把评价使用质量作为质量保证和质量控制一部分的情况;
- 供应:按合同约定向需方提供系统、软件产品或软件服务的契约行为,例如在合格性测试中确认质量的情况;
- 获取:包括从供方获取或采购系统、软件产品或软件服务时的产品选型和验收测试;
- 维护:基于使用质量测度的产品改善。

2 符合性

任何遵循本部分的质量需求规约或质量评价应:

- 选取如 GB/T 25000.10—2016 所定义的使用质量特性和/或子特性来进行指定或评价;
- 对于被选取的每项特性或子特性,宜全部使用第 8 章定义的通用(G)质量测度,如有不使用的测度,请阐述理由;
- 选取有关联的任何特定(S)质量测度;
- 如修改任何质量测度,请阐述每项修改的理由;
- 对要使用但却未包含在本部分内的附加质量测度加以定义;

f) 准确定义每个质量测度的可操作办法(比如所用测量方法或问卷的细节)。

注:在进行比较时使用相同测量方法很重要。

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 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)

GB/T 25000.1—2010 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第1部分:SQuaRE指南(ISO/IEC 25000—2005, IDT)

GB/T 25000.10—2016 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第10部分:系统与软件质量模型(ISO/IEC 25010:2011, MOD)

GB/T 25000.23—2019 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第23部分:系统与软件产品质量测量(ISO/IEC 25023:2016, MOD)

GB/T 25000.62—2014 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第62部分:易用性测试报告行业通用格式(CIF)(ISO/IEC 25062:2006, IDT)

GB/T 32261.2—2015 消费类产品和公用类产品的可用性 第2部分:总结性测试方法(ISO/TS 20282-2:2013, IDT)

ISO/IEC 25063 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 易用性通用行业格式(CIF):使用描述上下文环境[Systems and software engineering—Systems and software product Quality Requirements and Evaluation (SQuaRE)—Common Industry Format (CIF) for usability: Context of use description]

ISO/IEC 25064 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 易用性通用行业格式(CIF):用户需求报告[Systems and software engineering—Systems and software product Quality Requirements and Evaluation (SQuaRE)—Common Industry Format (CIF) for usability: User needs report]

ISO/IEC 31010 风险管理 风险评估技术(Risk management—Risk assessment techniques)

4 术语和定义

GB/T 25000.1—2010 和 GB/T 25000.10—2016 界定的以及下列术语和定义适用于本文件。

注:为便于使用,SQuaRE系列标准和其他ISO标准中的基本定义复制于此。

4.1

周境完备性 context completeness

在每个指定的使用周境中,产品或系统在所要求级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度。

注1:周境完备性是周境覆盖的子特性之一。

注2:改写GB/T 25000.10—2016,定义4.2.2.5.1。

4.2

周境覆盖 context coverage

在指定的使用周境和超出最初设定需求的周境中,产品或系统在有效性、效率、抗风险和满意度特性方面能够被使用的程度。

注:使用周境与使用质量和一些产品质量特性或子特性相关(即“指定条件”)。

[GB/T 25000.10—2016, 定义 4.2.2.5]

4.3

客户 customer

接受或使用某项产品或服务的组织或个人跟供方的关系。

注：本关系包括针对某项服务各种条款和条件的担保或协定。

4.4

有效性 effectiveness

用户实现指定目标的准确性和完备性。

[GB/T 18978.11—2004, 定义 3.2]

4.5

效率 efficiency

与用户实现目标的准确性和完备性相关的资源消耗。

注：相关的资源可包括完成任务的时间(人力资源)、原材料或使用的财务成本。

[GB/T 18978.11—2004, 定义 3.3]

4.6

使用周境 context of use

用户、任务、设备(硬件、软件和原材料)以及使用某系统、产品或服务的物理和社会环境。

注：改写 GB/T 18978.11—2004, 定义 3.5。

4.7

灵活性 flexibility

在超出最初设定需求的周境中,产品或系统在可接受级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度。

注 1：灵活性是周境覆盖的子特性之一。

注 2：改写 GB/T 25000.10—2016, 定义 4.2.2.5.2。

4.8

形成性评价 formative evaluation

为改善评价客体,尤其是仍在开发阶段的评价客体而设计和使用的评价。

[ISO/TS 18152:2010, 定义 4.6]

4.9

抗风险 freedom from risk

产品或系统的质量在经济现状、人的生命、健康或环境方面缓解或规避潜在风险的程度。

注 1：风险是一个给定威胁发生的概率及该威胁发生后的潜在不利后果的函数。

注 2：GB/T 25000 标准各部分考虑的风险是由产品质量缺陷引起的风险。

注 3：抗风险包括不利于用户、组织或项目的潜在风险的减免程度。

注 4：改写 GB/T 25000.10—2016, 定义 4.2.2.4。

4.10

目标 goal

预期结果。

[GB/T 18978.11—2004, 定义 3.8]

4.11

测度(名词) measure (noun)

以测量结果来赋值的变量。

注 1：术语“测度”作为集合名词,用来指代基本测度、导出测度和指标。

注 2：在本部分中,“测度”一词在被特性或子特性限定使用时指质量测度。

注 3：改写 ISO/IEC 15939:2007。

4.12

测量 measurement

以确定某一测度的值为目的的一组操作。

注 1: 测量可包括分配一个定性的类别,例如源程序的语言(ADA、C、COBOL 等)。

注 2: 改写 ISO/IEC 15939:2007。

4.13

测量函数 measurement function

为组合两个或多个质量测度元素而执行的算法或计算。

[GB/T 25000.21—2019,定义 4.7]

4.14

心理测量[学] psychometrics

用于开发有效和可靠的心理学测度相关理论和技术的研究领域。

4.15

使用质量 quality in use

指定用户使用产品或系统满足其要求的程度,以达到在指定的使用周境中的有效性、效率、满意度和抗风险等指定目标。

注 1: 软件产品或系统的使用质量能通过目标系统或软件产品的被实施系统的用户使用效果、实地测试效果或原型测试效果得到测量或评价。

注 2: 当使用质量被指定时,其意思是指定用户满足实现在指定使用周境中的有效性、效率、满意度和抗风险方面的指定目标的需要。

注 3: 改写 GB/T 25000.10—2016,定义 3.15。

4.16

质量测度 quality measure

至少两个质量测度元素值的测量函数的导出测度。

[GB/T 25000.21—2019,定义 4.13]

4.17

质量测度元素 quality measure element

以某一属性及其量化的测量方法(可选地包括通过数学函数所做的变换)来定义的测度。

[GB/T 25000.21—2019,定义 4.14]

4.18

质量模型 quality model

定义的特性集以及它们之间的关系集,为规约质量需求以及评价质量提供框架。

4.19

满意度 satisfaction

产品或系统在指定的使用周境中使用时,用户的要求被满足的程度。

注 1: 对于不直接与产品或系统交互的用户,仅与目标实现和可信性相关。

注 2: 满意度是指用户对其与产品或系统交互的反应,包括对产品使用的态度。

注 3: 用户包括:与系统交互以实现主要目标的主要用户;提供支持的辅助用户;以及接收输出但不与系统交互的间接用户。

注 4: 在本部分中,用户需要包括与产品、系统或服务使用相关联的用户愿望和期望。超越愿望和期望是显著增加满意度、改善用户体验的一种方法。

注 5: 改写 GB/T 25000.10—2016,定义 4.2.2.3。

4.20

利益相关方满意度 stakeholder satisfaction

产品或系统在指定的使用周境中使用时,利益相关方的要求被满足的程度。

注: 产品或系统的用户是利益相关方的类型之一,因此,用户满意度是利益相关方满意度的类型之一。

4.21

总结性评价 summative evaluation

为总结体现评价客体的优点或价值而设计的评价。

注 1: 评价结果能用于针对是否宜保留、更改或删除评价客体形成推荐性意见。

注 2: 设计一种兼具形成性和总结性的评价方法是可能的。

注 3: 采用总结性测试方法来实施总结性评价。

[GB/T 32261.2—2015, 定义 4.17]

4.22

系统 system

为达到一个或多个明确目的而组织起来的、相互作用的元素的组合体。

注 1: 一个系统可被认为是一个产品或它提供的服务。

注 2: 实际上, 对系统含义的解释通常通过使用一个联合名词来阐明, 如飞行器系统。或者, 单词“系统”可简单地由上下文相关的同义词来替代, 如飞行器, 虽然这可能使系统的观点不太明显。

注 3: 改写 ISO/IEC 15288:2015。

4.23

任务 task

实现目标所必需的活动集合。

注 1: 这些活动有可能是物理活动或认知活动。

注 2: 角色和职责可能决定目标和任务。

注 3: 改写 GB/T 18978.11—2004, 定义 3.9。

4.24

易用性 usability

产品或系统能够被指定用户为实现在指定周境中的有效性、效率和满意度指定目标而使用的程度。

注 1: 改编自 ISO 9241-210。

注 2: 易用性既可以从它的子特性角度当作产品质量特性来进行指定或测量, 也可以直接通过测度(使用质量的子集)来进行指定或测量。

[GB/T 25000.10—2016, 4.3.2.4]

4.25

使用差错 use error

引起有异于制造商意图或用户期望的系统响应结果的行为或行为缺失。

注: 改写 IEC 62366:2007。

4.26

用户 user

在系统使用过程中, 从系统中获益的个人或组织。

[ISO/IEC 15939:2007]

5 缩略语

下列缩略语适用于本文件。

QME: 质量测度元素(Quality Measure Element)

6 使用质量测度的使用

6.1 使用质量的应用

本部分为 GB/T 25000.10—2016 使用质量模型的如下特性和子特性提供了推荐的测度:

- 有效性；
- 效率；
- 满意度：
 - 有用性；
 - 可信性；
 - 愉悦性(用户体验)；
 - 舒适性(人体工效学)；
- 抗风险：
 - 经济风险缓解性；
 - 健康和安全风险缓解性；
 - 环境风险缓解性；
- 周境覆盖：
 - 周境完备性；
 - 灵活性。

使用质量是产品或系统为满足特定用户实现在特定使用周境中的有效性、效率、满意度和抗风险特定目标需要而能被使用的程度。

在使用质量模型中,术语“易用性”指由有效性、效率、满意度和周境覆盖所组成的使用质量子集。用户界面和交互的易用性和功能性测度在 GB/T 25000.23—2019 中给出。

低级别的产品易用性有可能引起低级别的易用性,进而引起不可接受的抗风险。不可接受的抗风险也可能由其他低级别的产品质量特性所引起。

使用质量测度测量用户与系统之间的交互结果。当某个已实现的系统或软件产品在用户环境中为预期目的而被使用时,该系统或软件产品的最终使用质量能得以测量。使用质量测度亦能用于概念设计和如下所示的开发阶段:

- 需求:使用质量需求能在设计和开发过程早期被加以指定,以便针对用户和其他利益相关方的质量体验期望设定高级别目标；
- 原型形成性评价:使用质量的特定问题能通过开发中对具有少量典型用户的原型进行测试来加以识别,期间,要利用产品易用性测度和非规范化的使用质量测度来识别问题、改善产品；
- 原型总结性评价:不同设计方案的最终使用质量估算结果能通过开发中对具有大量典型用户的原型进行测试来获得,期间,要利用规范化的使用质量测度,而且要估计由此产生的有效性、效率和满意度是如何充分缓解潜在风险的；
- 质量保证和控制过程:已实现系统的使用质量能针对质量需求被加以测试。

注 1: 附录 G 解释了不同 SQuaRE 质量模型之间的关系。

注 2: 在本部分中,“测度”(作名词使用)一词指质量测度。

6.2 使用质量的测量

使用质量不仅取决于软件或计算机系统的产品质量,还取决于产品被使用的具体周境(见 ISO/IEC 25063)。使用周境包括用户因素、任务因素以及能影响使用质量的物理和社会环境因素。因此,只有当测度产生在相同使用周境时,软件产品或系统的使用质量对比才有效。

这使得要预测或解读抗风险级别尤其困难,因为要控制可能影响抗风险的使用周境其他因素通常是不可能的。但是,要证明可能由低级别的易用性和产品质量所引起的潜在风险,以及要提出将缓解风险的易用性或产品质量的建议目标值,通常都是可能的(见图 2)。如果易用性或产品质量的任何测度都无法达到目标值,那么抗风险测量值的潜在影响就能被评估出来。

示例: 当设计航空公司订座系统时,针对用户预订某个期望日期、期望时间、期望目的地航班的行为设定一个高级

别的有效性(成功率)目标值,以使可能由任何差错所致的潜在经济后果的可能性降至最低。

许多测度是以可为特定需求定制的方式来定义的。因此,只有在测度以相同方式进入操作层面并且用于相同或足够相似的使用周境(即有相似类型的用户在相似环境中执行相似任务)的情况下,测度之间或者测度与目标值之间的对比才有意义。

通过观察典型用户在真实使用周境中执行典型任务,能评估有效性、效率和满意度(方法示例参见附录 D 和附录 E)。通过模拟真实使用环境(比如易用性实验室)或者观察产品操作使用,能获得测度。为了指定或测量使用质量,首先应识别预期使用周境的每个组件:用户、用户目标和使用环境。评价设计应与使用周境尽量密切匹配。同样重要的是,用户只能得到在操作环境中可用的帮助或辅助类型。当应用诸如 GB/T 32261.2—2015 所指定的方法时,这些测度能在保证过程中使用。

有些外部易用性测度(GB/T 25000.23—2019)被以相似方式测试,但是,还会在产品的更多一般使用过程中评价具体产品特征的使用,这是作为使用质量测试一部分的典型任务。

注 1: 附录 D 和 GB/T 32261.2—2015 提供了如何测量有效性、效率和满意度的示例。附录 C 提供了如何测量周境覆盖的示例。亦见参考文献[12]。

注 2: 附录 H 提供了关于 SQuaRE 质量测量概念的更多信息。

6.3 使用质量测度的解读

有些使用质量测度(例如任务完成时间)很难能孤立地解读。要能对使用质量测度进行比较以使其更易于解读,有以下几种比较方式:

- a) 一致性:将测度与特定业务或使用要求做比较(例如有可能在十分钟内完成任务);
- b) 基准:将测度与有相同用途的相同或相似产品或系统的基准做比较(例如在新系统完成任务的时间有可能等于或少于在旧系统所用的时间);
- c) 时序:对一段时间的趋势做比较(例如系统每个新版本原型的用户差错减少量);
- d) 熟练度:用从训练有素或专家型用户使用中获取的值做比较(例如与熟练用户相比,新用户所花时间长多少);
- e) 满意度人群常模:当有前值数据库时,测度能表达为先前评分值不低于现值的用户的百分率。

第 8 章展示的使用质量测度包括了以上述方式之一规范化的示例,但是,大部分测度可能有几种规范化形式,如附录 D 所示。

未规范化测度(如差错数或任务用时)能用于识别单一用户或小用户群所遇到的特定问题。

要获得可靠的有效性、效率或满意度测度,应从足够多的用户处获取数据,这些用户执行若干任务来确保所达成的目标值在统计置信度上达到期望的级别。

注: GB/T 32261.2—2015 提供了关于大众用途产品有效性、效率和满意度总结性测试的方法示例。

6.4 适用使用质量测度的选取

能影响特定使用质量测度选取的因素包括:

- 有效性、效率、满意度和抗风险的相对重要性;
- 可能对经济状况、生命、健康或环境造成风险的有效性、效率或满意度特定方面;
- 应用某个测度所必备的技能 and 知识。

有关使用质量评价过程的更多信息参见附录 F。

6.5 其他标准中对使用质量各方面的描述

以下标准提供了有关使用质量特定方面的更多信息:

- GB/T 18978.11—2004: 有效性、效率和满意度的定义在 GB/T 18978.11—2004 中与在 GB/T 25000.10—2016 中相似,因此,在 GB/T 25000.22 中的有效性、效率和满意度的使用质

- 量测度亦能作为 GB/T 18978.11—2004 中所定义的易用性测度(参见附录 E);
- GB/T 25000.62—2014:指定了关于易用性的总结性测试结果的文档记录方法;
- GB/T 32261.2—2015:提供了关于测量消费品的有效性、效率和满意度的缜密方法论;
- ISO/IEC 25063:解释了在测量易用性时应识别的使用周境元素;
- ISO/IEC 25064:解释了宜记录(可能包括使用质量需要的)用户需要的方法。

7 质量测度记录格式

以下是第 8 章表格中每个测度都具备的信息:

- a) ID:指质量测度的标识码。每个 ID 均由以下三部分构成:
 - 质量特性和子特性代码。分别用“Ef”和“Ey”格式代表有效性和效率,用“ABc”格式代表满意度、抗风险和周境覆盖,其中“A”代表质量特性、“Bc”代表质量子特性;
 - 质量子特性中连续排列的序号;
 - 代表普遍适用、可能用于广泛情况的“G”,或代表专用于特定需要的“S”。
- b) 名称:质量测度的名称。
- c) 描述:质量测度所提供的信息。
- d) 测量函数:展示各个质量测度元素如何组合来计算出质量测度的公式。
- e) 方法:能用于获取测度的方法种类:
 - 用户表现测量法:有效性和效率测度(参见 F.3.3.2 和 F.3.3.3);
 - 用户行为测量法:收集用户动作数据;
 - 数据自动收集法:借助软件去收集用户动作数据;
 - 问卷调查法:满意度测度(参见 F.3.3.5);
 - 业务分析法:分析业务活动及其结果;
 - 软件与易用性分析法:分析由人为或系统错误所引起的潜在风险;
 - 使用统计法:分析由先前人为或系统错误所导致的结果;
 - 周境描述分析法:为评估预期使用质量而分析使用周境;
 - 检测法:为识别潜在问题而对系统进行检测。

8 使用质量测度

8.1 概述

本章按质量特性和子特性列出质量测度,并沿用 GB/T 25000.10—2016 中的排列顺序:有效性、效率、满意度、抗风险和周境覆盖。

有效性、效率和满意度的测度值的获得取决于使用周境(见 6.2),因此,应结合任何测量结果对测度为之产生(或将要为之产生)的用户类型、任务类型和环境类型加以明述。

标准和规章符合性要求能被识别为系统需求的一部分,但是这超出了质量模型的范围。

注 1:本质量测度清单并非最终版本,有可能在本部分的未来版本中加以修订。诚邀本部分读者提供反馈。

注 2:关于用于几个质量测度的一些质量测度元素的定义,参见附录 I。

8.2 有效性测度

有效性测度评估用户特定目标实现的准确性和完备性(见表 1)。

注 1:有效性测度不考虑目标如何实现,只考虑目标实现的程度(参见 F.3.1.2)。

注 2:“由用户造成的差错”一语指用户没有执行预期动作。这些差错有时被称为“使用差错”,以强调差错主因可

能在于系统设计不良。

表 1 有效性测度

ID	名称	描述	测量函数	方法
Ef-1-G	任务完成率	在没有协助的情况下正确完成的任务比例	$X=A/B$ A = 已完成的单一任务数量 B = 所尝试的单一任务总数	用户表现测量法
注 1: 本测度能够针对一个用户或用户组进行测量。 注 2: 若任务能够部分完成,则目标实现率测度更加适用。 注 3: 若任务具有不同的复杂度,则可在公式中使用加权任务:$X=\sum(i=1\cdots n)W_i\times A_i/B$,其中,$i$ 指任务序号,W_i 代表任务难度,W_i 总和为 1.0。 注 4: 本测度可应用于需求中所识别出的任务,或者应用于用户尝试的任务。				
Ef-2-S	目标实现率	在没有协助的情况下正确实现的任务目标比例	$\{X=1-\sum A_i X\geq 0\}$ A_i = 每个缺失或错误目标在任务输出中的比值(最大值=1)	用户表现测量法
注: 对于每个潜在的缺失或不完整的组件都根据其对企业或用户输出值的减损程度给予一个权重 A_i。(若权重总和超过 1,则质量测度通常被设为 0,尽管这可能预示着负面结果和过度风险。)将评分方案应用于一系列任务输出,并不断调整权重,直至获得可重复、可复制且有意义的结果,从而迭代改进该评分方案。 示例: 与若干潜在客户对日志和通讯录的潜在差错的业务影响进行讨论,得出下列计算平均目标实现率的评分方案(以百分比表示): ——安装:成功安装所有组件计为 100%,漏装必要子组件每个扣 20%; ——新增联系人:正确填入所有信息计为 100%,缺填信息每项扣 50%,填错信息字段每项扣 20%,打错字每个扣 5%; ——新增会议:正确填入所有信息计为 100%,填错时间或日期为 0%,填错信息字段每项扣 20%,打错字每个扣 5%。 扣分组合大于或等于 100%时,目标实现率计为 0%。				
Ef-3-G	任务中的差错数	用户在某一任务中造成的差错数量	$X=A$ A = 用户在某一任务中造成的差错数量	用户表现测量法
注 1: 由用户造成的差错数可包含全部差错,或仅包含未修正的差错,或仅包含导致该任务未正确完成的差错。 注 2: 差错计数测度可用于比较相同任务在不同境况下执行的异同,例如比较某一开发中系统的不同版本的异同。 注 3: 要比较不同任务的差错,可将差错数与每个任务的动作数进行关联。 注 4: 只有当差错具有同等重要性或者被加权时,才适合对其进行比较。 注 5: 用户问题矩阵可用于对差错进行分析,指出有多少用户在何种组合中出了哪种问题。				
Ef-4-G	出错任务率	出现用户所造成差错的任务比例	$X=A/B$ A = 出错任务数量 B = 任务总数	用户表现测量法
注: Ef-3-G 的注释适用于此。				
Ef-5-G	任务差错密度	造成差错的用户的比例	$X=A/B$ A = 造成差错的用户数量 B = 执行任务的用户总数	用户表现测量法
注: Ef-3-G 的注释适用于此。				

8.3 效率测度

效率测度评估与用户目标实现的准确性和完备性相关的资源消耗(见表2)。

注1: 最常见资源是完成任务所用的时间,其他相关资源可能包括用户人力、材料或使用财务成本。

注2: 效率测度最典型的是与使用不同产品或版本时或者不使用产品时的效率做比较。此外,也可与专家效率做比较。

表2 效率测度

ID	名称	描述	测量函数	方法
Ey-1-G	任务用时	成功完成一项任务所花费的时间	$X = T$ $T = \text{任务用时}$	用户表现测量法
注: 通过对某项任务的常规用户用时与专家用时以及重做任务的用时做比较,能测量易学性(见 GB/T 25000.23)。				
Ey-2-S	时间效率	用户在使用系统时实现目标的效率	$X = A / T$ $A = \text{实现的目标数量}$ $T = \text{时间}$	用户表现测量法
注1: 时间效率是衡量生产力的指标,即单位时间内实现的目标数量。效率随有效性的增加及任务用时的减少而提高。这使得效率能被比较,例如对快速而易错的界面与慢速而容易的界面之间的比较。 注2: 若 Ey-1-G 任务完成率已被测量,则时间效率能被测量为任务完成率与时间之比。这测量了单位时间内成功完成任务的比例。测量值越高,表示在短时间内成功任务比例越高。 注3: 能够将时间效率跟专家时间效率,或针对不同产品或版本的时间效率,或手动完成任务的时间效率做比较。 注4: 若实现的目标存在不同的值,则能够对其加权。				
Ey-3-S	成本效率	用户的成本效率	$X = A / B$ $A = \text{执行任务的总成本}$ $B = \text{实现的目标数量}$	用户表现测量法
注1: 实现的目标的示例包括既定的商业目标、必要的信息检索或者系统输出。 注2: 若实现的目标存在不同的值,则能够对其加权。 注3: 成本可能包括诸如用户时间成本、他人协助时间成本以及计算资源、电话呼叫和材料等方面的成本。 注4: 本测度可能要规范化以避免产生系统空置成本。				
Ey-4-S	生产性时间比	用户执行生产性动作的时间比例	$X = T_a / T_b$ $T_a = \text{生产性时间} = \text{任务完成时间} - \text{求助时间} - \text{差错修复时间} - \text{无效搜索时间}$ $T_b = \text{任务用时}$	用户表现测量法
注: 通过对交互视频的分析,能够识别在求助、差错修复和无效搜索方面所花费的非生产性时间。				
Ey-5-S	非必要动作率	对达成任务无必要的由用户执行的动作比例	$X = A / B$ $A = \text{对达成任务无实际必要的动作数量}$ $B = \text{用户执行动作的数量}$	用户表现测量法或数据自动收集法
注1: 本测度对通过做若干选择(例如用鼠标、触摸或语音发出指令)来完成任务的情况最为有用。对于更复杂的任务而言,能够将动作定义为包含数据输入。 注2: 本测度的目的与生产性时间比测度的目的相近,但更易于测量。				

表 2（续）

ID	名称	描述	测量函数	方法
Ey-6-S	疲劳影响	在持续使用后的人力表现的衰减	$X=1-A/B$ A=当前表现 B=初始表现	用户表现测量法或数据自动收集法
注 1：适用于由有经验的用户持续使用的情况。 注 2：表现指有效性或效率的任何适用测度（在必要时进行规范化，使其数值越大越好）。 注 3：生理性测度能用于评估疲劳所带来的影响。 注 4：个人疲劳的评估能够通过问卷调查法进行测量。 注 5：疲劳测度仅适用于执行重复性任务的经验丰富的用户。 注 6：计算机系统和工作实践能被设计用于减轻疲劳。 注 7：越接近 0 越好。				

8.4 满意度测度

8.4.1 概要

满意度测度评估产品或系统在指定的使用周境中使用时，用户的要求被满足的程度（见表 3）。满意度测度的范围可能是：

- a) 满意度的总体通用测度（见 SUs-1-G）；
- b) 满意度特定子特性的测度（见 8.4.2、8.4.3、8.4.4 和 8.4.5）；
- c) 由个体子特性测度相组合所产生的满意度总测度。

注：个体子特性组合测度能按其对总体满意度的相对重要性和贡献予以加权。

用户通过在有两个刻度点（例如“同意”和“不同意”）或多个刻度点（例如由“非常同意”过渡至“非常不同意”）的量表上选取某个值来对满意度问卷的问题做出应答。

满意度问卷总评分范围通常在 0 分～100 分。某些问卷（如系统易用性量表（SUS），见参考文献 [14]）有历史结果数据库，使得先前给类似产品或系统打过相同或更高分数的那部分用户的分数能转换成某个值。

本条提供了能用于获取满意度不同方面的质量测度的方法示例。使用现成的公开问卷能获得特定测度。

注 1：用户包括提供支持的辅助用户和接收输出的间接用户，他们皆不与系统进行交互。

注 2：满意度是用户对其与产品或系统交互所做的回应，包括对产品使用的态度。

注 3：用户满意度受用户对软件或计算机系统属性（如由外部测度所测量的属性）的感知以及用户对使用中的有效性、效率、抗风险的感知所影响。

注 4：用户满意度能在开发期间作为易用性测试的一部分进行测量，以了解如何改进设计。在产品发布后以问卷形式对用户满意度进行测量，以帮助用户了解该产品是否符合他们的需求。

注 5：心理测量量表公认是可靠的、有效的。使用包含规范化心理测量量表的问卷的好处在参考文献 [15] 中有解释。

表 3 通用满意度测度

ID	名称	描述	测量函数	方法
SUs-1-G	总体满意度	用户的总体满意度	$X = \sum A_i$ A_i = 对某提问的回应	问卷调查法
注：关于满意度总体测度的举例有“净推荐值” ^[11] 和“单项难易度问卷” ^[13] 。				

8.4.2 有用性测度

有用性测度评估用户在感知上对实用目标(包括使用结果和使用后果)实现的满意程度(见表 4)。

表 4 有用性测度

ID	名称	描述	测量函数	方法
SUs-2-G	特征满意度	用户对特定系统特征的满意度	$X = \sum A_i$ A_i = 对与特定特征相关提问的回应	问卷调查法
注：这是一种采用李克特量表 ^a 的典型非确认问卷。在将问卷项目进行组合以得出一个总分时,能视不同提问的重要程度予以加权。				
SUs-3-G	自主使用率	选择使用某系统或功能的潜在用户比例	$X = A/B$ A = 使用特定功能、应用或系统的用户数量 B = 可能使用特定功能、应用或系统的潜在用户数量	用户行为测量法或数据自动收集法
注：本测度能用于可能要识别某个功能、应用或系统在何时达到合适使用率的时候,例如在监视用户行为采样时。				
SUs-4-G	特征利用率	已识别的系统用户集里使用某具体特征的用户比例	$X = A/B$ A = 使用某具体特征的用户数量 B = 系统已识别用户集的用户数量	用户行为测量法或数据自动收集法
注 1：特征能从单一功能到系统子集的不同粒度等级上进行定义。 注 2：分值低可能说明该特征是无用的,或仅对某用户子集适用,或用户并未理解如何使用,或用户不知道该特征的存在。				
SUs-5-G	用户投诉率	发出投诉的用户比例	$X = A/B$ A = 投诉的用户数量 B = 系统的用户数量	用户行为测量法
SUs-6-G	具体特征用户投诉率	针对某具体特征的用户投诉比例	$X = A/B$ A = 针对某具体特征的用户投诉数量 B = 针对所有特征的用户投诉总数	用户行为测量法
^a 对于某类调查提问的应答范围。其中,答题者要对照一个量表来为自己对某段陈述的反应进行评分。该量表典型表现为由正面评分依次过渡至负面评分,中性评分居中。				

8.4.3 可信性测度

可信性测度评估用户或其他利益相关方对产品或系统表现如其所愿的信心程度(见表 5)。

表 5 可信性测度

ID	名称	描述	测量函数	方法
STr-1-G	用户可信性	用户信任系统的程度	$X=A$ A = 从可信性问卷得出的心理 测量值	问卷调查法
注：有关可信性问卷的示例，见参考文献[9]。				

8.4.4 (用户体验)愉悦性测度

愉悦性测度评估用户愉悦感需要得到满足的程度(见表 6)。

注 1：改自 GB/T 25000.10—2016 中的定义：“用户因个人要求被满足而获得愉悦感的程度”。

注 2：用户需要包括对获得新知识和新技能、传递人格同一性、唤起愉快回忆以及参与交互活动等方面的需要。

表 6 (用户体验)愉悦性测度

ID	名称	描述	测量函数	方法
SPl-1-G	用户愉悦性	与同类系统的使用愉悦感平均水平相比，用户获得愉悦感的程度	$X=A$ A = 从愉悦性问卷得出的心理 测量值	问卷调查法
注：有关愉悦性心理测量问卷的示例，参见参考文献[7]和[16]。				

8.4.5 (人体工效学)舒适性测度

舒适性测度评估用户身体舒适感需要得到满足的程度(见表 7)。

注：改自 GB/T 25000.10—2016 中的定义：“用户生理上感到舒适的程度”。

表 7 (人体工效学)舒适性测度

ID	名称	描述	测量函数	方法
SCo-1-G	身体舒适性	与同类系统的使用舒适感的平均水平相比，用户获得舒适感的程度	$X=A$ A = 从舒适性问卷得出的心理 测量值	问卷调查法
注 1：身体舒适性能受用户使用计算机系统时不得不采取的姿势或动作以及受系统使用环境所影响。 注 2：有关舒适性心理测量问卷的示例，参见参考文献[8]。				

8.5 抗风险测度

8.5.1 概要

抗风险测度评估产品或系统的质量缓解或规避不利于用户、组织或项目的潜在风险(包括不利于经济状况、生命、健康或环境的风险)的程度。

注：改自 GB/T 25000.10—2016 中的定义：“产品或系统在经济现状、人的生命、健康或环境方面缓解潜在风险的程度。”

非预期后果风险的根源在于任何产品质量特性有缺陷(见 ISO/IEC 25063)或有效性、效率、满意度、周境覆盖级别不够。

非预期后果风险能对下列类型的利益相关方造成影响：

- a) 产品或系统用户：
 - 在使用产品或系统时的健康和安全风险；
 - 未能实现预期结果的不良后果。
- b) 产品或系统使用组织：
 - 因易用性差致使组织犯错而造成的组织名誉损失和财务损失；
 - 因操作安全缺失或对安全或隐私保护不足所引发的风险。
- c) 产品或系统开发组织：
 - 因系统、产品或服务的设计和开发无法生产出符合预期质量目标的系统所引发的经济后果的风险；
 - 因质量缺陷导致经济或名誉受损进而致使产品或系统无人购买或服务无人使用的风险。
- d) 更广泛群体：
 - 健康和安全后果或负面环境影响的风险。

表 8 展示了何种利益相关方能受何种风险后果的影响。

表 8 不同类型利益相关方的不良后果类型

利益相关方	后果类型		
	健康和安全	经济状况	环境
产品或系统用户	健康和安全风险缓解性测度 因易用性差所致的压力	未能实现预期结果的后果	
产品或系统使用组织		经济风险缓解性测度	
产品或系统的开发或获取组织		经济风险缓解性测度	
更广泛群体	健康和安全后果的风险		经济风险缓解性测度

8.5 所定义的质量测度能以不同方式进入操作层面,以切合利益相关方的使用需要。对于表达为积极结果或消极结果的测度,风险缓解指规避不可接受的值。

注：通过指定可接受的最低质量级别并以此作为评估准则,能控制不良后果风险。指定较高的质量级别可能要求投入额外资源,但能为改善经济、健康或环境结果提供机会。见图 2。



图 2 与质量级别关联的风险和机会

注 1：风险是某个既定威胁发生的概率与该威胁发生的潜在不良后果之间的函数。

注 2：由质量对结果影响所引发的潜在风险通常以经济状况、生命、健康或环境风险级别评估为基础，该级别由（预测或）已测量的产品质量（功能性、性能效率、兼容性、易用性、可靠性、信息安全性、维护性或可移植性）所产生。

有关风险评估的信息，见 ISO/IEC 31010。

8.5.2 经济风险缓解性测度

经济风险缓解性测度评估质量对风险与机会并存的财务状况、效率运营、商业资产、名誉等资源相关经济目标的影响（见表 9）。

注 1：风险缓解性用于缓解可能因产品质量差所引发的不可接受经济结果风险。

注 2：根据历史数据的实际值能估计这些测度的期望值。

注 3：测量结果很难能够解读，因为这些测度的促成因素有很多，很难能够将产品或系统质量对每个经济结果实际测度的促成作用孤立看待。这些测度的主要价值在于根据质量如何影响经济结果的先前经验去估计质量对经济目标的影响。跟系统/软件产品行为（如 REc-5-S、REc-6-S、REc-7-S、REc-8-G）密切相关的财务测度将更易于解读。

表 9 经济风险缓解性测度

ID	名称	描述	测量函数	方法
REc-1-G	投资回报率	在投资上的回报	$X = (A - B) / B$ A = 获得的额外收益 B = 投资额	业务分析法
注：获得的收益可包括如下示例：减少的人事开支、缩减的库存资产、减持的股份以及由集中采购所降低的材料成本。				
REc-2-G	投资回报用时	为实现期望投资回报率所花费的时间	$X = T$ T = 实现投资回报率的时间	业务分析法
注：本测度能与某个可接受的投资回报率实现时间做比较。				
REc-3-G	业务绩效	与目标对比的盈利能力或销售额	$X = A_s / A_t$ A = 公司盈利能力或销售额 (a = 实际值, t = 目标值)	业务分析法
注：本测度能与另一家公司的信息技术投资额或销售额做比较。				

表 9 (续)

ID	名称	描述	测量函数	方法
REc-4-G	信息技术投资收益(率)	信息技术投资收益测度(例如使用平衡计分卡获得)的实际值与目标值之比	$X = A_s / A_t$ A = 信息技术投资收益测度 (a = 实际值, t = 目标值)	业务分析法
注: 平衡计分卡从财务、客户、业务操作过程和人力资源发展等视角来评价信息技术投资收益。这些收益的任何一种都能成为某个测度的基础。				
REc-5-S	客户服务表现	对预期客户服务水平的实现程度	$X = A / B$ A = 实际服务水平 B = 预期服务水平	业务分析法
注: 若服务水平超出预期, 该值能大于 1。 示例: 交付延迟程度、客户服务获取的平均等待时间。				
REc-6-S	网站访客客户转化率	成为客户的具体网页访客比例	$X = A / B$ A = 成为客户的访客数量 B = 具体网页的单一访客数量	业务分析法
注: 本测度能基于特定网页或整个网站进行测量。				
REc-7-S	单客收入	来自单个客户的收入	$X = A$ A = 单个客户收入	业务分析法
注: 客户有若干种属性, 如“现有的”和“新的”。这些属性能用于评价错失新功能提供机会的状况。				
REc-8-G	伴经济后果差错率	存在伴随着经济后果的人为或系统差错的使用情况比例	$X = A / B$ A = 伴随经济后果的差错数量 B = 使用情况总数	业务分析法、软件与易用性分析法
注 1: 差错能按经济重要度加权。使用情况能由交易或时间界定。差错能包括数据损坏。 注 2: 经济后果能对于某组织或某个人用户而言。 示例: 某人未能有效且有效率地实现预期交通票购买目标的经济后果可能导致不能如期出行或需要重新购票。				

8.5.3 健康和安全风险缓解性测度

健康和安全风险缓解性测度评估质量对健康和安全目标的影响(如避免重复性劳损)(见表 10)。

注 1: 有关诸如安全性或保密性质量特性的特定测量, 参见关于医疗设备的 IEC 62366 以及由 IEC 65 和 ISO/IEC JTC 1/SC 27 所提供的国际标准。

注 2: 风险缓解性可用于缓解由于产品质量差而可能在健康和安全方面产生不可接受测度值的风险。

表 10 健康和安全风险缓解性测度

ID	名称	描述	测量函数	方法
RHe-1-G	用户健康报告 (发生)率	报告因使用产品导致健康问题的用户比例	$X = A/B$ A = 报告健康问题的用户数量 B = 用户总数	使用统计法
注 1: 健康问题能包括重复性劳损、疲劳、头痛等。 注 2: 本测度可能按使用时长予以加权。				
RHe-2-G	用户健康和安 全影响	产品对用户健康和安全的 影响	$X = \frac{1}{T_b} \sum_{i=1}^n (T_{a_i} \times S_i)$ n = 受影响的人数 T_{a_i} = 第 i 个人受影响的时长 S_i = 第 i 个人受影响的严重程度 T_b = 系统启动以来的运行 时长	使用统计法
注: 影响能包括心理上和身体上的健康和安全的,例如因系统用户界面难用所造成的压力。				
RHe-3-G	使用系统对人 身安全的影响	因使用系统而受影响的人员 所受危险的发生率	$X = A/B$ A = 处于危险中的人数 B = 使用系统而可能受影响的 总人数	使用统计法
注 1: 本测度的示例之一是“对患者人身安全的影响”,其中 A 为治疗不当患者人数,B 为患者总数。 注 2: 对于某些用途,相对比值而言,受影响人数可能是关联性更高的测度。				

8.5.4 环境风险缓解性测度

环境风险缓解性测度评估质量对环境目标的影响(见表 11)。

注: 风险缓解性用于缓解因产品质量差所导致的不可接受环境后果风险。

表 11 环境风险缓解性测度

ID	名称	描述	测量函数	方法
REn-1-G	环境影响度	产品或系统的制造和使用对 环境影响的实际值与目标值 之比	$X = A_a/A_t$ A = 环境影响 (a = 实际值, t = 目标值)	使用统计法
注 1: 环境影响可能包括污染、噪声或全球变暖等取决于系统类型的后果。 注 2: 环境风险可能由软件缺陷或由用户界面设计差所致的用户差错风险所造成。				

8.6 周境覆盖测度

8.6.1 概要

周境覆盖测度评估在指定的使用周境和超出最初设定需求的周境中,产品或系统在有效性、效率、抗风险和满意度特性方面能够被使用的程度。

- 注 1: 附录 C 提供了周境覆盖测度示例。
- 注 2: 不同等周境指因其中用户、任务和/或环境不同而导致易用性存在重大差异的周境(见 GB/T 32261.2—2015)。

8.6.2 周境完备性测度

周境完备性测度用于评估在每个指定的使用周境中,产品或系统在要求级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度(见表 12)。

- 注 1: 这是对 GB/T 25000.10—2016 所述定义“在所有指定的使用周境中,产品或系统在有效性、效率、抗风险和满意度特性方面能够被使用的程度”的修正版本。
- 注 2: 周境完备性能够根据指定用户为实现所有预期使用周境中指定有效性、效率、抗风险和满意度目标的真实使用情况,或根据支持在所有预期使用周境中使用的产品属性呈现情况来加以指定或测量。

表 12 周境完备性测度

ID	名称	描述	测度函数	方法
CCm-1-G	周境完备性	产品或系统能够在其中以可接受级别的易用性和风险被使用的预期使用周境比例	$X = A / B$ $A =$ 伴随可接受易用性和风险级别的周境数量 $B =$ 必需而不同的使用周境总数	用户表现分析法或周境描述分析法
注 1: 通过分析或用户测试,能评估某产品或系统是否对所有预期用户、任务和环境类型组合都具有可接受级别的易用性。 注 2: 不同使用周境能够按其重要度加权。 示例 1: 要求是让系统在所有预期使用周境中都具有足够好的使用质量。 示例 2: 系统仅在四个预期使用周境中的三个具有足够好的使用质量。				

8.6.3 灵活性测度

灵活性测度用于评估在超出最初设定需求的周境中,产品或系统在可接受级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度(见表 13)。

- 注 1: 灵活性使产品能够对很可能无法提前预知的境况、机会和个人偏好加以考虑。
- 注 2: 通过分析产品特性和使用周境来评估产品能够被额外类型用户为实现在额外类型使用周境中的有效性、效率、抗风险和满意度额外类型目标而使用的程度,或通过测试在所述额外使用周境中有用户的产品,又或通过产品对新型用户、任务和环境的适应调整能力及其如 ISO 9241-110 所定义的个性化适配能力,能够对灵活性进行测量。

表 13 灵活性测度

ID	名称	描述	测度函数	方法
CFI-1-S	使用周境灵活性	产品在额外使用周境(不同类型的用户、任务和环境)中无需修改或仅作简单修改就能使用的程度	$X = A / B$ A = 产品能够以可接受使用质量被使用于其中的额外周境数量 B = 产品可能被使用于其中的额外周境总数	用户表现分析法或周境描述分析法
注 1: 简单修改意思是产品能由用户自定义设置或仅仅是文本和/或数据要修改。 注 2: 越接近 1 越好。 示例: 最初为在某具体市场使用而设计的某产品有可能会在其他广泛场景(B)里使用,但当前设计只在这些场景的某子集(A)里是可用的。				
CFI-2-S	产品灵活性	产品能够被修改以满足额外的用户需求的容易程度	$X = \frac{1}{B} \sum_{i=1}^n A_i$ A _i = 对第 i 个需求的易修改性(如 GB/T 25000.23 所指定) B = 指定用户的新需求总数	用户表现分析或检测
CFI-3-S	熟知度独立性	产品能够由无特定知识、技能或经验的人所使用的程度	$X = A / B$ A = 无特定知识、技能或经验但却能使用产品的额外用户组数量 B = 无特定知识、技能或经验的潜在用户组总数	用户表现分析或检测
注 1: 产品可能最初预期供具备特定知识、技能或经验的某用户组使用,但潜在地可由更广泛类型用户使用。 注 2: 若产品仅能由预期用户组使用,则测度值为 0。若产品能由任何人使用,则测度值为 1。				

附 录 A
(资料性附录)

本部分与 ISO/IEC 25022:2016 相比的结构变化情况

本部分与 ISO/IEC 25022:2016 相比在结构上有少量调整,具体章条编号对照情况见表 A.1。

表 A.1 本部分与 ISO/IEC 25022:2016 的章条编号对照情况

本部分章条编号	对应的 ISO/IEC 25022:2016 章条编号
附录 A	—
附录 B	—
附录 C	附录 A
附录 D	附录 B
附录 E	附录 C
附录 F	附录 D
附录 G	附录 E
附录 H	附录 F
附录 I	附录 G

附 录 B
(资料性附录)

本部分与 ISO/IEC 25022:2016 的技术性差异及其原因

表 B.1 给出了本部分与 ISO/IEC 25022:2016 的技术性差异及其原因。

表 B.1 本部分与 ISO/IEC 25022:2016 的技术性差异及其原因

本部分章条编号	技术性差异	原因
3	关于规范性引用文件,本部分做了具有技术性差异的调整,调整的情况集中反映在第3章“规范性引用文件”中,具体调整如下: ——用修改采用国际标准的 GB/T 25000.10—2016 代替了 ISO/IEC 25010(见第1章、第2章); ——增加引用了 GB/T 18978.11—2004、GB/T 25000.1—2010、GB/T 25000.23—2019、GB/T 32261.2—2015、GB/T 25000.62—2014、ISO/IEC 25063、ISO/IEC 25064、ISO/IEC 31010	所列出的文件在本部分的正文中被引用
4.23	修改注2中的“工作职责”为“角色和职责”	使本部分与 GB/T 25000.23—2019 在表述上保持一致
7	修改列项 e) 中九种方法的中文表达句式为“……法”	在表达上更符合中文习惯,避免在下文引述时产生歧义
8.1	增加了正文中对附录 I“用于定义质量测度的质量测度元素”的引述文字“注2:关于用于几个质量测度的一些质量测度元素的定义,参见附录 I。”	按照我国标准的编写规则
8.5.3	修改表10中 RHe-3-G 测度的名称“Safety of……”(……的安全(率))的中文用词为“……安全的影响”	ISO/IEC 25022:2016 中,该名称的英文字面意思跟相关描述和测量函数所表达的真实意思相反
8.5.4	修改表11中 RE _n -1-G 测度的名称“Environmental impact”(环境影响)的中文用词为“环境影响度”	ISO/IEC 25022:2016 中,该测度的名称和质量测度元素 A 使用相同的英文短语,在中文表达上应予以明确区分,避免产生歧义
附录 H	修改了图 H.1 中表示质量测度元素的框图	使本部分与 25000 标准的其他部分保持一致

附 录 C
(资料性附录)
周境覆盖测量方法示例

C.1 概要

本附录提供了关于周境覆盖测量方法的示例。

周境覆盖被定义为在指定的使用周境和超出最初设定需求的周境中,产品或系统在有效性、效率、抗风险和满意度特性方面能够被使用的程度。周境覆盖由子特性周境完备性和灵活性所组成。

周境完备性被定义为在每个指定的使用周境中,产品或系统在要求级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度。

C.2 某航空公司网站的周境完备性示例

某航空公司订票网站的指定使用周境可能包括该系统应对用户类型(U)、任务(T)和网页浏览器(B)的下列所有组合而言都是可用的:

——指定用户类型:

- U1:频飞旅客;
- U2:偶飞旅客(有该网站使用经历)。

——指定任务:

- T1:航班选择;
- T2:支付;
- 其他任务。

——指定环境:网页浏览器:

- B1:IE 浏览器;
- B2:火狐浏览器;
- 其他浏览器。

——用户类型、任务和环境的组合(每种组合分别定义一个特定使用周境,应在其中达到指定级别的有效性、效率、满意度和抗风险等使用质量):

- U1 T1 B1;
- U1 T1 B2;
- U1 T2 B1;
- 其他。

为每个使用周境所指定的使用质量级别可能不同,但设计目标是对于所有指定浏览器和用户类型而言的使用质量是相同的。

对周境完备性要求进行指定,目的是确保在所有指定使用周境中均达到质量需求。

C.3 某航空公司网站的周境完备性评价

如果对现有系统使用周境的每个排列组合都开展用户测试,则代价昂贵。替代办法之一是,对系统进行检测,如果没有发现预期会引起支付任务中不同浏览器或不同用户类型间使用质量差异的可识别

问题,则只需要测量下列排列组合:

- U1 T1;
- U2 T1;
- U1+U2(结合) T2。

C.4 某航空公司网站的灵活性测度

灵活性被定义为在超出最初设定需求的周境中,产品或系统在可接受级别的有效性、效率、抗风险和满意度特性方面能够被使用的程度。

灵活性设计目标之一是利用已知能够兼容最广范围浏览器的技术去实现网站。

灵活性测量能根据该实现技术对不同浏览器的兼容性评估或者通过使用一系列替代浏览器去测试该网站而开展。

对灵活性要求进行指定,目的是在使用现有和将有浏览器访问网站时更有可能达到可接受级别的有效性、效率和满意度。

附 录 D
(资料性附录)
使用质量测度规范化

表 D.1 展示了可应用于每个质量测度的规范化选项(如 6.3 所述)。

表 D.1 规范化选项

			规范化选项				
ID	名称	测量函数	一致性	基准	时序	熟练度	人群常模
Ef-1-G	任务完成率	$X = A / B$ $A =$ 已完成的单一任务数量 $B =$ 所尝试的单一任务总数	✓	✓	✓	✓	
Ef-2-S	目标实现率	$\{X = 1 - \sum A_i \mid X \geq 0\}$ $A_i =$ 每个缺失或错误目标在任务输出中的比值(最大值=1)	✓	✓	✓	✓	
Ef-3-G	任务中的差错数	$X = A$ $A =$ 用户在某一任务中造成的差错数量	✓	✓	✓	✓	
Ef-4-G	出错任务率	$X = A / B$ $A =$ 出错任务数量 $B =$ 任务总数	✓	✓	✓	✓	
Ef-5-G	任务差错密度	$X = A / B$ $A =$ 造成差错的用户数量 $B =$ 执行任务的用户总数	✓	✓	✓	✓	
Ey-1-G	任务用时	$X = T$ $T =$ 任务用时	✓	✓	✓	✓	
Ey-2-S	时间效率	$X = A / T$ $A =$ 实现的目标数量 $T =$ 时间	✓	✓	✓	✓	
Ey-3-S	成本效率	$X = A / B$ $A =$ 执行任务的总成本 $B =$ 实现的目标数量	✓	✓	✓	✓	
Ey-4-S	生产性时间比	$X = T_a / T_b$ $T_a =$ 生产性时间 = 任务完成时间 - 求助时间 - 差错修复时间 - 无效搜索时间 $T_b =$ 任务用时	✓	✓	✓	✓	
Ey-5-S	非必要动作率	$X = A / B$ $A =$ 对达成任务无实际必要的动作数量 $B =$ 用户执行动作的数量	✓	✓	✓	✓	

表 D.1 (续)

ID	名称	测量函数	规范化选项				
			一致性	基准	时序	熟练度	人群常模
Ey-6-S	疲劳影响	$X=1-A/B$ A = 当前表现 B = 初始表现	✓	✓	✓	✓	
SUs-1-G	总体满意度	$X=\sum A_i$ A_i = 对某提问的回应	✓	✓	✓	✓	✓
SUs-2-G	特征满意度	$X=\sum A_i$ A_i = 对与特定特征相关提问的回应	✓	✓	✓	✓	
SUs-3-G	自主使用率	$X=A/B$ A = 使用特定功能、应用或系统的用户数量 B = 可能使用特定功能、应用或系统的潜在用户数量	✓	✓	✓	✓	
SUs-4-G	特征利用率	$X=A/B$ A = 使用某具体特征的用户数量 B = 系统已识别用户集的用户数量	✓	✓	✓	✓	
SUs-5-G	用户投诉率	$X=A/B$ A = 投诉的用户数量 B = 系统的用户数量	✓	✓	✓		
SUs-6-G	具体特征用户投诉率	$X=A/B$ A = 针对某具体特征的用户投诉数量 B = 针对所有特征的用户投诉总数	✓	✓	✓		
STr-1-G	用户可信性	$X=A$ A = 从可信性问卷得出的心理测量值	✓	✓	✓	✓	✓
SPl-1-G	用户愉悦性	$X=A$ A = 从愉悦性问卷得出的心理测量值	✓	✓	✓	✓	✓
SCo-1-G	身体舒适性	$X=A$ A = 从舒适性问卷得出的心理测量值	✓	✓	✓	✓	✓
REc-1-G	投资回报率	$X=(A-B)/B$ A = 获得的额外收益 B = 投资额	✓	✓	✓		
REc-2-G	投资回报用时	$X=T$ T = 实现投资回报率的时间	✓	✓	✓		
REc-3-G	业务绩效	$X=A_s/A_t$ A = 公司盈利能力或销售额 (a = 实际值, t = 目标值)	✓	✓	✓		

表 D.1 (续)

ID	名称	测量函数	规范化选项				
			一致性	基准	时序	熟练度	人群常模
REc-4-G	信息技术投资收益(率)	$X = A_s / A_t$ A = 信息技术投资收益测度 (a = 实际值, t = 目标值)	✓	✓	✓		
REc-5-S	客户服务表现	$X = A / B$ A = 实际服务水平 B = 预期服务水平	✓	✓	✓		
REc-6-S	网站访客客户转化率	$X = A / B$ A = 成为客户的访客数量 B = 具体网页的单一访客数量	✓	✓	✓		
REc-7-S	单客收入	$X = A$ A = 单个客户收入	✓	✓	✓		
REc-8-G	伴经济后果差错率	$X = A / B$ A = 伴随经济后果的差错数量 B = 使用情况总数	✓	✓	✓		
RHe-1-G	用户健康报告(发生)率	$X = A / B$ A = 报告健康问题的用户数量 B = 用户总数	✓	✓	✓		
RHe-2-G	用户健康和安全影响	$X = \frac{1}{T_b} \sum_{i=1}^n (T_{a_i} \times S_i)$ n = 受影响的人数 T_{a_i} = 第 i 个人受影响的时长 S_i = 第 i 个人受影响的严重程度 T_b = 系统启动以来的运行时长	✓	✓	✓		
RHe-3-G	使用系统对人身安全的影响	$X = A / B$ A = 处于危险中的人数 B = 使用系统而可能受影响的总人数	✓	✓	✓		
REn-1-G	环境影响度	$X = A_s / A_t$ A = 环境影响 (a = 实际值, t = 目标值)	✓	✓	✓		
CCm-1-G	周境完备性	$X = A / B$ A = 伴随可接受易用性和风险级别的周境数量 B = 必需而不同的使用周境总数	✓	✓	✓		

表 D.1 (续)

ID	名称	测量函数	规范化选项				
			一致性	基准	时序	熟练度	人群常模
CFl-1-S	使用周境灵活性	$X = A / B$ $A =$ 产品能够以可接受使用质量被使用于其中的额外周境数量 $B =$ 产品可能被使用于其中的额外周境总数	✓	✓	✓		
CFl-2-S	产品灵活性	$X = \frac{1}{B} \sum_{i=1}^B A_i$ $A_i =$ 对第 i 个需求的易修改性 (如 GB/T 25000.23 所指定) $B =$ 指定用户的新需求总数	✓	✓	✓		
CFl-3-S	熟知度独立性	$X = A / B$ $A =$ 无特定知识、技能或经验但却能使用产品的额外用户组数量 $B =$ 无特定知识、技能或经验的潜在用户组总数	✓	✓	✓		

附录 E
(资料性附录)

使用本部分测量 GB/T 18978.11—2004 中的易用性

GB/T 18978.11—2004 和 GB/T 25000.10—2016 中的易用性、有效性和效率的定义基本相同，而满意度的定义则不同。表 E.1 列出了本部分和 GB/T 18978.11—2004 中的术语定义。

注：在 GB/T 18978.11—2004 中，有效性、效率和满意度是易用性的子特性。

表 E.1 本部分和 GB/T 18978.11—2004 中的术语定义

	GB/T 25000.10—2016	GB/T 18978.11—2004
易用性	在指定的使用周境中，产品或系统在有效性、效率和满意度特性方面为了指定的目标可为指定用户使用的程度	在指定的使用周境中，系统、产品或服务在有效性、效率和满意度特性方面为了指定的目标可为指定用户使用的程度
有效性	用户实现指定目标的准确性和完备性	用户实现指定目标的准确性和完备性
效率	与用户实现目标的准确性和完备性相关的资源消耗	与用户实现目标的准确性和完备性相关的资源消耗
满意度	产品或系统在指定的使用周境中使用时，用户的要求被满足的程度	对使用产品的正面态度和无不适感

因此，8.2(有效性)与 8.3(效率)中的所有质量测度均能用于测量 GB/T 18978.11—2004 中的有效性与效率。能用于测量 GB/T 18978.11—2004 中满意度的质量测度分别是来自 8.4.1(概要)、8.4.2(有用性)和 8.4.3(可信性)的正面态度测度以及来自 8.4.5(舒适性)的无不适感测度。

附 录 F
(资料性附录)
使用质量评价过程

F.1 概要

本附录中的总体过程如下：

- 确立评价需求：
 - 明确评价目的；
 - 界定产品类型；
 - 规定质量模型；
- 规定评价：
 - 界定使用周境；
 - 选择评价周境；
 - 选取测度；
 - 确定评估准则；
 - 解读测度；
- 设计评价；
- 执行评价：
 - 开展用户测试并收集数据；
- 编制报告。

注 1：本附录的章条结构大致遵循 GB/T 25000.40 中描述的评价过程结构。

注 2：GB/T 32261.2—2015 提供了关于消费品和公共用品的有效性、生产效率和满意度评价过程的详细规格说明，该评价过程与本附录提到的评价过程一致。

F.2 确立评价需求

F.2.1 明确评价目的

评价使用质量的目的在于评估产品能满足用户需要、在特定使用周境内实现特定目标的程度。

F.2.1.1 获取

在开发工作之前，组织在力图获取适应自身需求的具体产品时，能够用使用质量作为框架来规定使用质量需求，这些要求宜被产品满足，且能够接受验收测试。宜识别待测量使用质量的所在周境，宜选取有效性、效率、满意度和抗风险的测度，且宜确立基于这些测度的验收准则。

F.2.1.2 供给

供方能够评价使用质量，以确保产品满足各种特定类型用户和使用环境的需要。向潜在需方提供使用质量结果，将有助于需方判断产品是否满足其特定需要（见 GB/T 25000.62—2014 的示例）。

F.2.1.3 开发

清晰理解不同使用场景里的用户使用质量需求，将有助于开发团队在做设计决策时以满足用户真

实需要为导向,并将开发目标聚焦在满足使用质量准则上。在开发结束时,这些准则能被评价。

F.2.1.4 运行

通过对使用质量各方面的测量,组织在运行一个系统时能评价该系统满足其需要的程度,并能评估未来版本可能被要求做出何种改变。

F.2.1.5 维护

对于软件维护人员来说,维护任务的使用质量能被测量。

F.2.1.6 软件移植

对于软件移植人员来说,移植任务的使用质量能被测量。

F.2.2 识别产品类型

评价使用质量应有工作原型或成品。

F.2.3 规定质量模型

使用的质量模型是 GB/T 25000.10—2016 给出的使用质量模型。使用质量定义为指定用户使用产品或系统满足其要求的程度,以达到在指定的使用周境中的有效性、效率、满意度和抗风险等指定目标。

F.3 规定评价

F.3.1 界定使用周境

F.3.1.1 概述

为了规定或测量使用质量,应界定使用周境的各个组件:用户、用户目标和使用环境。通常情况下,不可能去测试所有可能的使用周境,所以应选取重要或典型的用户组和任务。

F.3.1.2 用户

应规定能影响用户使用产品表现的用户特征。这些特征可能包括:知识、技能、经验、教育水平、训练水平、身体素质以及运动和感官能力。对于不同类型的用户,例如具备不同经验水平或执行不同角色任务的用户,应加以定义。

注:在人口老龄化的情况下,年长和行动障碍人群的特征能对其表现产生影响。这可能涉及:使之表现更佳的附加经验;或者使之表现存在差异或面临更困难使用周境的体质、运动或感官方面的能力。

F.3.1.3 目标(预期结果)

应规定产品使用目标。目标规定的是要实现什么,而不是如何实现。目标能分解为若干子目标,子目标规定整体目标的组件以及满足该子目标的准则。举个例子,如果目标是为了完成一张客户订单表格,子目标可能是在每个区域输入正确的信息。整体目标的宽度取决于评价范围。实现目标的必要活动是任务。

F.3.1.4 环境

操作环境

由于操作环境会影响到软件表现,所以应规定硬件和软件的操作环境。操作环境包括网络响应时间等更广泛的方面。

用户环境

还应规定能影响用户表现的使用环境的所有方面,诸如物理环境(如工作场所、家具)、周遭环境(如气温、光线)以及社会文化环境(如工作实践、他人帮助、激励)。

F.3.2 选择评价周境

用于评价的周境尽可能匹配产品实际使用的一个或多个环境,这点很重要。为预测产品被真实使用时达到的使用质量级别,要获取若干测度,这些测度的有效性取决于用户、任务和环境代表真实情景的典型程度。在一个极端情况下,人们能在“现场”做测量,基于真实情景来评价产品的使用质量。情况的另一个极端是,人们能在“实验室”里评价产品的个别方面,当中,使用周境的各个方面彼此关联,并被典型、可控地重塑。采用实验室方法的好处是,有机会更加严格地控制预计会严重影响已达成使用质量级别的变量,从而能更加精确地测量。坏处是,人造的实验室环境可能产生不切实际的结果。

F.3.3 选取测度

F.3.3.1 测度的选择

要规定或评价使用质量,通常应至少测量一次有效性、效率、满意度和抗风险(若合适)方面的测度。

测度及其受测周境的选择取决于测量参与方的目标。每个测度跟目标的相对重要度都宜加以考虑。举个例子,在使用不频繁时,可理解性和易学性测度很可能比使用质量测度更加重要。

使用质量测度宜以反映用户和产品交互结果的数据为基础。可能利用一些客观手段,如对输出结果、工作速度或个别事件进行测量,去收集数据。要不然,能从用户感受、信仰、态度或偏好等主观反映来收集数据。客观测度直接标示有效性和效率,而主观测度能与满意度直接关联。

评价能在现场和实验室环境设定一体沿线的不同点上开展,这些设定取决于要调查的问题和可测试产品的完备性。测试环境和测度的选择将取决于测量活动目标以及这些目标与设计周期的关系。

F.3.3.2 有效性

有效性测度测量能达成目标的准确性和完备性。

例如,如果期望目标是准确地重新制作一份 2 页的特定格式文档,那么能通过拼写错误数量和格式偏差数量来规定或测量准确性,并且能通过把转抄文档字数除以源文档字数来规定或测量完备性。

F.3.3.3 效率

效率测度把已达成有效性级别和资源消耗联系在一起。相关资源能包括精神或体力付出、时间、物资材料或金钱成本。例如,能用有效性除以人力付出来测量人力效率,用有效性除以时间来测量时间效率,或者用有效性除以成本来测量经济效率。

如果期望目标是打印一份报告,那么能用报告打印件可用份数除以任务资源消耗量来规定或测量效率,后者包括被消耗的工时、工序和材料(含必要返工成本)等。

F.3.3.4 抗风险

抗风险测度跟软件或计算机系统超时操作风险、使用状况和使用周境有关。能根据操作风险削减和意外风险削减来分析抗风险。操作风险削减是软件在正常运行时满足客户要求而不伤害其他资源和环境的能力。意外风险削减是软件在正常范围之外运行和转移资源以防止风险扩散的能力。

F.3.3.5 满意度

满意度测量用户摆脱不适感的程度和他们的产品使用态度。

满意度能通过若干主观量表来规定和测量,如产品喜爱程度、产品使用满意度、多任务执行时工作负荷承受能力、个别使用质量目标(如效率或易学性)达成程度等。满意度的其他测度可能包括使用中记录的正面和负面评论数量。额外信息能获自一些较长期的测度,如缺勤率、关于用户认知或体力工作负荷过多还是过少的观察结果、健康问题报告、用户要求更换工作的频率,等等。

满意度的主观测度通过对用户主观表达的反应、态度或观点的强度进行量化来获得。量化过程能通过多种方式完成,例如,让用户给出跟他们在某个时刻感觉强度相对应的一个数字,让用户按偏好程度给产品排序,或者使用基于问卷的态度量表。

开发得当的态度量表能快速使用,公认可靠,且不必掌握特殊应用技能。利用心理测量技术开发的 attitude 问卷,将提供可靠性和有效性的公认、量化评估结果,能避免诸如造假、正面或负面回应偏见、社会期望等因素的影响,还使这些结果能跟历史回应的已确立规范进行比较。参见参考文献以了解关于测量计算机系统满意度的问卷示例。

F.3.4 确定评估准则

使用质量测度准则值的选择取决于产品要求和组织在设定准则时的需要。使用质量目标能跟某一初始目标(如敲入一个字母)或子目标(如搜索和替换)相关联。将使用质量目标焦点放在最重要用户的目标上,意味着忽视很多功能,但却很可能是最实际的方法。为特定子目标设定使用质量目标,能使评价较早地进入开发过程。

当设定一组用户的准则值时,该准则能设为某个平均值(例如任务完成平均时间不超过 10 min),该值针对个人(例如所有用户能在 10 min 内完成任务),或针对一定比例用户(例如 90% 用户能在 10 min 内完成任务)。

当设定准则时,宜注意给每个测量项分配合适的权重。举个例子,要设定基于差错的准则,很可能应分配不同权重来反映不同类别差错的相对重要程度。

F.3.5 解读测度

由于使用质量特征的相对重要程度取决于使用周境和使用质量规定或评价目的,因此,没有关于宜如何选择或组合测度的普遍性规则。

在将任何使用质量测度结果推广到另一个周境时,后者在用户类型、任务或环境方面都会明显不同,宜加以关注。如果使用质量测度是在短时间内获得的,那么这些值可能并未考虑到会严重影响使用质量的罕见事件,例如间歇性系统错误。

对于一般用途的产品来说,通常应规定或测量多个不同、典型周境的使用质量,其结果将是所涉及周境和所能完成任务的子集。这些周境的使用质量能互不相同。

F.4 设计评价

评价宜在尽可能接近产品使用状况的情况下开展。重要的是:

——用户是产品用户人群的典型;

注: GB/T 32261.2—2015 的附录 C 详细解释了如何选取典型用户样本,包括如何对少数用户群体(如特定民族或特定障碍人群)加以考虑。

——任务是系统预期完成任务的典型;

——状况是产品正常使用状况(包括帮助获得途径、时间压力和干扰因素)的典型。

通过控制评价周境,经验表明利用仅有 8 名参与者的样本就能获得可靠结果(见 GB/T 25000.62—2014)。

F.5 执行评价

当评估使用质量时,重要的是用户独立工作,仅可获得在正常使用状况下可用的帮助形式。测量有效性、效率和满意度也是一样,通常用文档记录用户遇到的问题,并在该时段结束时与用户讨论辩明这些问题。用视频记录评价常常很有用,这样可以进行更详细的分析和制作视频剪辑。在视频远程监视下,用户也更容易不受打扰地工作。

F.6 制作报告

如果要制作综合报告,可从《行业通用格式》(GB/T 25000.62—2014)得到针对有效性、效率和满意度等使用质量组件的良好报告框架。

附录 G
(资料性附录)
不同质量模型之间的关系

系统/软件产品质量是指其满足各利益相关方明确的或隐含的要求进而提供价值的程度。这些明确的或隐含的要求在 SQuaRE 标准中以质量模型的方式提出,质量模型将系统/软件产品的质量按特性分类,有些情况下被细分为子特性。

用户对质量的要求包括在指定使用周境中的系统使用质量需求。当使用系统使用质量特性和子特性来规定质量的外部测度和内部测度时,这些已确定的要求可被利用。

软件质量可通过测量内部属性(通常为中间产品的静态测量)或通过测量外部属性(通常为测量代码执行时的行为)来评价。系统产品的使用质量可通过测量使用质量的属性(当产品在实际或模拟使用时)来评价。软件的适当的内部属性是达到要求的外部行为的先决条件,而适当的外部行为则是达到使用质量的先决条件(如图 G.1 所示)。

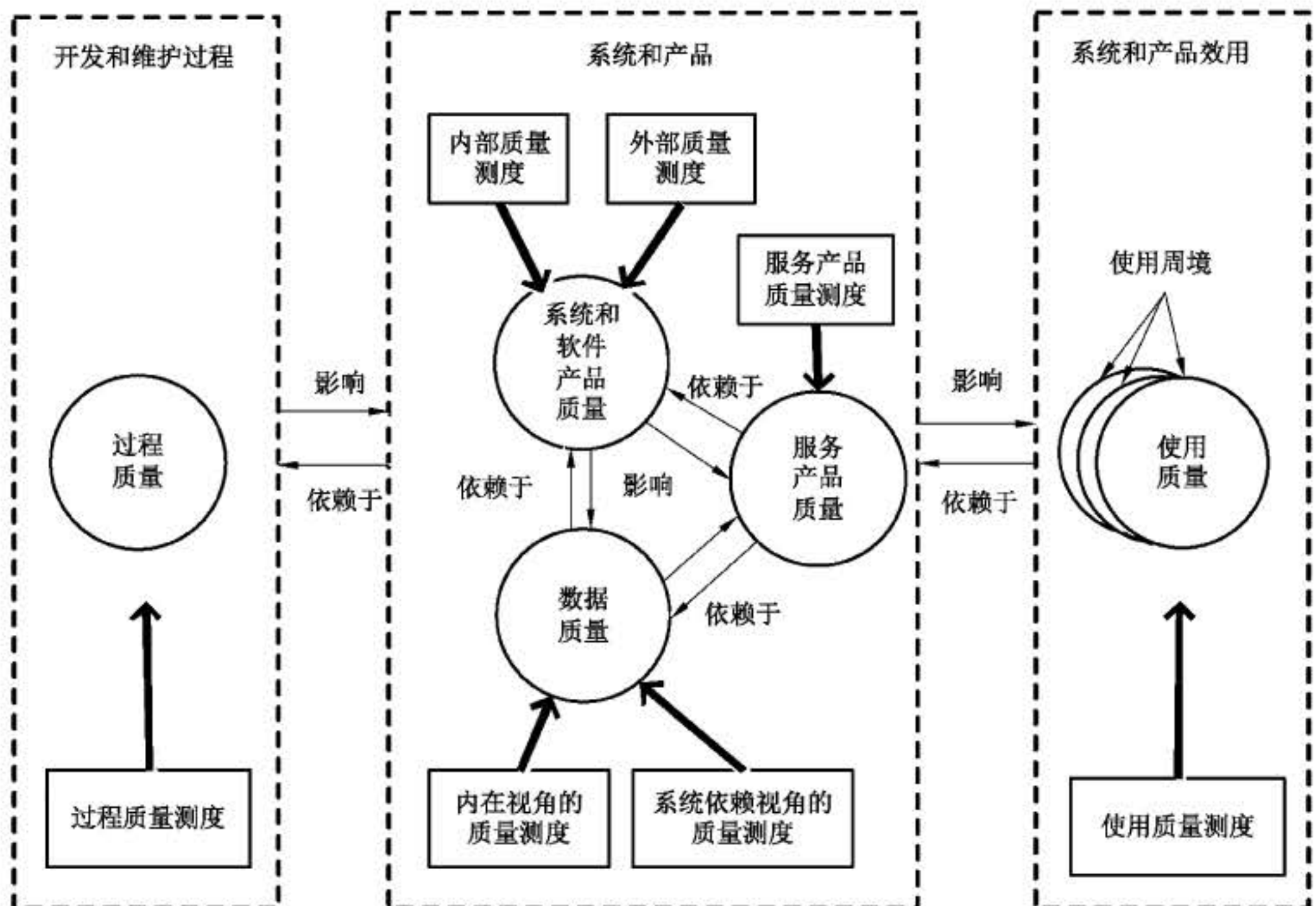
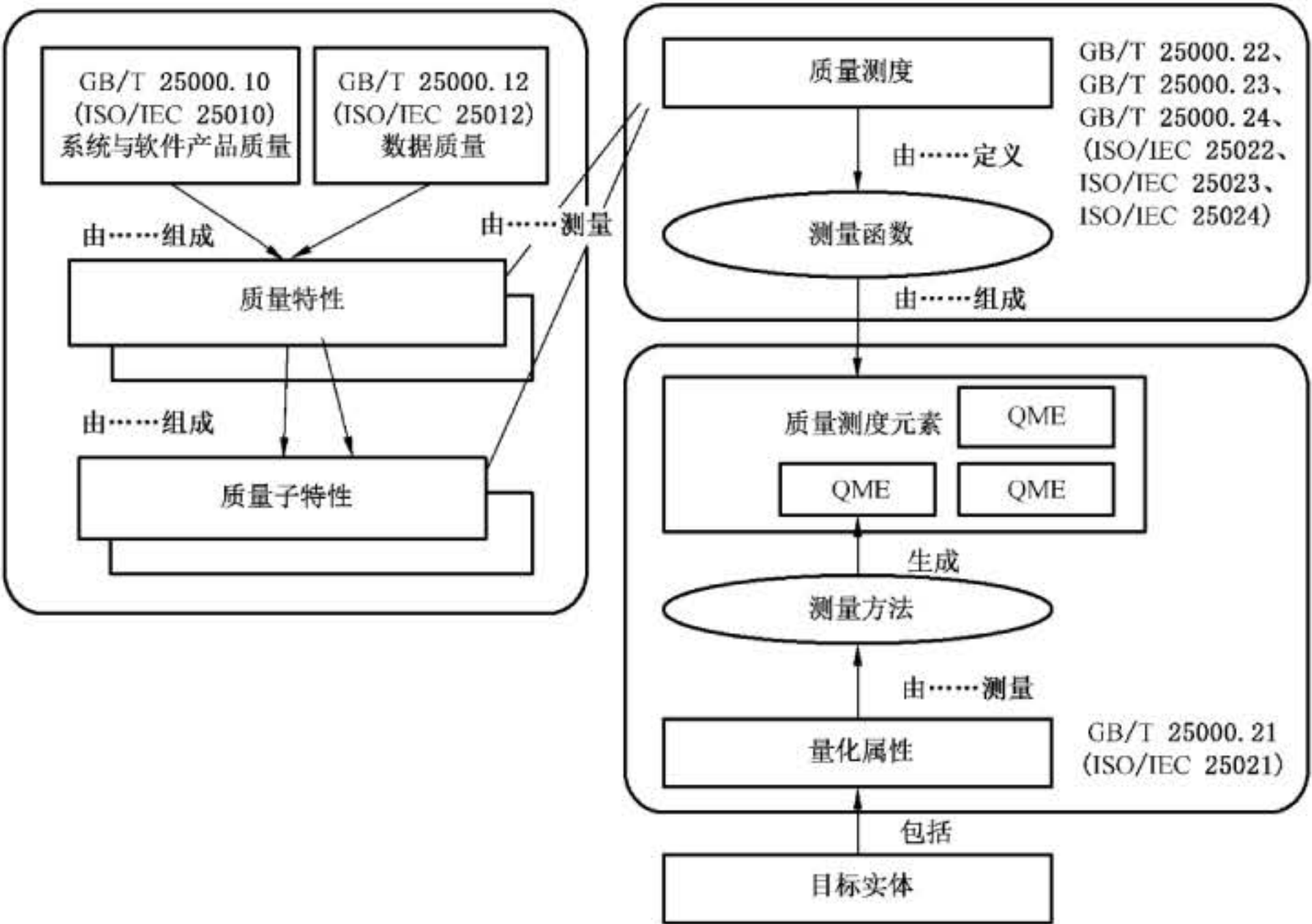


图 G.1 质量测度类型之间的关系

附录 H
(资料性附录)
质量测量概念

一个系统/软件产品可测量的、与质量相关的属性被称为与质量测度相关联的量化属性。这些属性可通过测量方法来测量。测量方法是一种逻辑操作序列,用来量化关于特定标度的属性。测量方法的应用结果被称为质量测度元素。

质量特性和子特性可以通过测量函数来量化。测量函数是用于组合质量测度元素的算法。测量函数的应用结果被称为质量测度。这样说来,质量测度就是对质量特性和子特性的量化。多个质量测度可用于测量质量特性或子特性(如图 H.1 所示)。



注：目标实体可以是系统、软件产品、数据或用户。参见 GB/T 25000.10—2016,图 5。

图 H.1 质量模型、质量测度(QM)、质量测度元素(QME)、量化属性和目标实体间的关系

附 录 I

(资料性附录)

用于定义质量测度的质量测度元素

大部分质量测度元素在每个质量测度的定义中都有充分描述。以下给出用于几个质量测度的一些质量测度元素的定义：

——任务：

用户为达到某目标而进行的活动。

——目标：

任务目的。

——用户差错：

用户无意中偏离正确方向与实际的行为或想法。

——出错任务：

用户执行过程中无意偏离正确方案的任务。

参 考 文 献

- [1] GB/T 25000.21—2019 系统与软件工程 系统与软件质量要求和评价(SQure) 第21部分:质量测度元素(ISO/IEC 25021:2012, IDT)
 - [2] ISO/IEC 12207:2008 Systems and software engineering—Software life cycle processes
 - [3] ISO/IEC 15288:2015 Systems and software engineering—System life cycle processes
 - [4] ISO/IEC 15939:2007 Systems and software engineering—Measurement process
 - [5] IEC 62366:2007 Medical devices—Application of usability engineering to medical devices
 - [6] ISO 9241-110:2006 Ergonomics of human-system interaction—Part 110: Dialogue principles
 - [7] HASSENZAHN M., & MONKA. The inference of perceived usability from beauty. Hum. Comput. Interact 2010, 25(3) pp. 235-260
 - [8] HERNANDEZ L. (2002). Evaluation of different scales for measurement of perceived physical strain during performance of manual tasks. Int. J. Occup. Saf. Ergon. 2002, 8(4) pp. 413-432
 - [9] JIAN J.-Y., BISANTZ A.M., DRURY G. Foundations for an empirically determined scale of trust in automated systems. Int. J. Cogn. Ergon. 2000, 4(1) pp. 53-71
 - [10] MINISTRY OF ECONOMY, Trade and Industry, Japan (METI) (2011) A Set of Metrics for Information Systems/Software product quality in Japan. http://www.meti.go.jp/policy/monoinfo/service/joho/cloud/2011/11_05.pdf
 - [11] NET PROMOTER SCORE (2014). http://en.wikipedia.org/wiki/Net_Promoter
 - [12] SAURO J., & LEWIS J. Quantifying the User Experience. Morgan Kaufmann, 2010
 - [13] Sauro, J. (2010) Single Ease Question (2010). www.measuringu.com/blog/single-question.php
 - [14] SAURO J. (2011). Measuring Usability With The System Usability Scale (SUS). <http://www.measuringusability.com/sus.php>
 - [15] SAURO J. (2012) Advantages Of Standardized Usability Questionnaires. <http://www.measuringusability.com/blog/standardized-usability.php>
 - [16] WATSON D., CLARK L.A., TELLEGEN A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 1988, 54(6) pp. 1063-1070
-

中 华 人 民 共 和 国
国 家 标 准
系统与软件工程
系统与软件质量要求和评价(SQuaRE)
第 22 部分:使用质量测量
GB/T 25000.22—2019

*

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.org.cn

服务热线:400-168-0010

2019 年 8 月第一版

*

书号:155066·1-63001

版权专有 侵权必究



GB/T 25000.22-2019