

Windows 11系统可用性评估与用户满意度分析

葛 纓^{1*}, 季嘉尧^{1*}, 刘 杰^{2#}, 庞惠珊³

¹重庆文理学院, 重庆市情绪与心理健康重点实验室, 暨用户体验与人机交互技术联合研究所, 重庆

²重庆第二师范学院教师教育学院, 暨西部儿童脑与认知科学研究院, 重庆

³西南交通大学交通运输与物流学院, 四川 成都

收稿日期: 2025年6月29日; 录用日期: 2025年7月23日; 发布日期: 2025年7月31日

摘 要

Windows 11系统作为WIN系列最新操作系统, 在带来新体验的同时, 也引发了用户各类使用问题。目的: 验证国际经典量表CSUQ、ASQ、QUIS在国内的适用性, 探究Windows 11系统的可用性和用户满意度现状。方法: 通过收集量表数据检验三个量表的国内适用性, 采用量表法获取用户的使用现状并进行分析。结果: CSUQ、ASQ、QUIS三个量表均表现出良好的本土适用性(高低分组差异显著、球形检验显著、内部一致性高、效标关联效度显著等); 用户对Windows 11系统总体满意度较高, 不同性别在Windows 11系统的可用性和满意度评价上不存在显著性差异、职业与专业在Windows 11系统的可用性和满意度评价上具有交互效应、不同电脑类型在Windows 11系统的可用性和满意度评价上存在显著差异。结论: 国际经典量表CUSQ、ASQ、QUIS在国内具有较高的文化适用性, 其中不同性别在评价上不存在显著性差异、职业与专业在评价上具有交互效应、不同电脑类型在评价上存在显著差异。

关键词

Windows 11系统, 用户体验, 系统可用性, 满意度分析

Usability Evaluation and User Satisfaction Analysis of Windows 11 System

Ying Ge^{1*}, Jiayao Ji^{1*}, Jie Liu^{2#}, Huishan Pang³

¹User Experience & Human-Computer Interaction Technology Institute, Key Laboratory of Emotion and Mental Health in Chongqing, Chongqing University of Arts and Sciences, Chongqing

²West China Institute of Children's Brain and Cognition, School of Teacher Development, Chongqing University of Education, Chongqing

³School of Transportation and Logistics, Southwest Jiaotong University, Chengdu Sichuan

Received: Jun. 29th, 2025; accepted: Jul. 23rd, 2025; published: Jul. 31st, 2025

*共同第一作者。

#通讯作者。

文章引用: 葛纓, 季嘉尧, 刘杰, 庞惠珊. Windows 11 系统可用性评估与用户满意度分析[J]. 交叉科学快报, 2025, 9(4): 582-600. DOI: 10.12677/isl.2025.94073

Abstract

As the latest operating system in the Windows series, Windows 11 has introduced new user experiences while also giving rise to various user issues during usage. Purpose: To validate the applicability of internationally recognized scales CSUQ, ASQ, and QUIS in the Chinese context, and to explore the current status of Windows 11's usability and user satisfaction. Method: The applicability of the three scales in the domestic context was examined by collecting scale data. The scale method was employed to obtain and analyze users' current usage status. Results: All three scales (CSUQ, ASQ, and QUIS) demonstrated good local applicability (e.g., significant differences between high and low score groups, significant sphericity test results, high internal consistency, and significant criterion-related validity). Users showed relatively high overall satisfaction with Windows 11. No significant differences were found between genders in evaluations of Windows 11's usability and satisfaction; there was an interaction effect between occupation and major in these evaluations; and significant differences were observed across different computer types in such evaluations. Conclusion: The internationally classical scales CSUQ, ASQ, and QUIS exhibit high cultural applicability in the Chinese context. Specifically, no significant gender differences exist in evaluations; occupation and major have an interaction effect on evaluations; and significant differences are present across different computer types in evaluations.

Keywords

Windows 11 System, User Experience, System Usability, Satisfaction Analysis

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 导言

2021年6月24日微软正式发布 Windows 操作系统最新版本 Windows 11, 这为全球的计算机用户带来了新奇而独特的体验, 其中该版本的开始菜单栏更加直观, 提供更加便捷的导航和搜索体验, 任务栏外观更加现代而简洁, 功能更加丰富, 这些新的设计不仅考虑到了审美吸引力, 并且保持了所有应用之间的统一性。

但随着操作系统布局和外观的改变, 适应新界面给不熟悉 Windows 11 系统的用户带来挑战, 旧的软件或硬件还会出现兼容性问题, 需要用户自己来更新驱动程序以获得更加丰富而稳定的功能。为了帮助用户适应这些改变, 熟悉新的操作系统和功能, 向 Windows 11 系统用户收集操作的反馈建议, 提升 Windows 11 的系统可用性和用户满意度至关重要。

1.1. 可用性相关研究

可用性概念起源于人机交互和软件心理学领域, 自上世纪七十年代提出以来, 已成为计算机软件评估的关键标准, 并逐渐扩展到网站、移动应用等多个领域[1]。ISO 9241-11 国际标准对可用性的定义是: 产品、服务与系统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、效率和用户主观满意度[2]。可用性包括以下六个方面的具体属性: 易学, 用户可以在短时间内学会并开始使用系统来做某些事情; 有效, 用户完成特定任务和达到特定目标时所具有的正确和完整程度; 高效, 当用户学会使用系统后, 可以具有高的生产力水平; 好记, 系统应当被容易记忆, 学习中断后不用一切操作从头学起; 少

错, 用户在使用系统的过程中能够少出错, 并且在出错之后能够迅速恢复; 满意度, 用户在使用时主观上感到满意, 喜欢使用系统。

在实际运用中, 可用性测试的方法有很多, 其中以测量法最为常用。可用性测量的使用范围不仅涵盖了产品的整体评估、任务评估和网站感知可用性评估等方面[3], 还可以根据特定领域和行业的需求进行定制和应用, 在学术研究和实践中发挥着重要作用, 帮助生产者提升产品或服务的用户体验。通过可用性测量, 可以及时发现产品中的潜在问题并加以改进, 提升产品的可用性。例如在 Windows 系统设计中可用性量表就可以帮助在产品更新升级阶段收集用户反馈, 优化产品设计和功能。

在国内使用可用性量表进行测量的研究中, 系统可用性量表[4] (System Usability Scale, SUS)是可用性评估中广泛使用的标准化问卷, 占工业可用性研究中问卷使用的 43%。还有一些与系统可用性主题紧密相关的量表, 如计算机系统可用性问卷[5] (Computer System Usability Questionnaire, CSUQ) 3.0 目前尚未在国内进行深入的本土化验证, CSUQ3.0 由苏珊娜·亨利在 20 世纪 80 年代末领导修订, alpha 系数为 0.95, 主要用于在任务结束后对计算机系统总体进行评估, 广泛应用于各种计算机系统的可用性测试, 包括桌面应用程序、网站、移动应用等, 通过收集用户在使用系统后的反馈, 帮助开发人员识别系统中的可用性问题, 从而进行改进和优化。

1.2. 用户满意度相关研究

用户满意度最早是管理学中的重要概念, 是指用户在购买商家提供的商品及服务时是否达到或者超出预期程度的一种心理状态[6], 通过测试用户对产品和服务的满意度, 从而得出产品和服务质量的评价数值, 反映产品和服务的质量状况。具体而言, 它代表用户对于产品或服务满足其需求程度的感知和评价, 是一个量化的指数, 用于衡量产品或服务在市场上的表现和用户接受度, 用户满意度的高低则直接影响到企业的声誉、客户忠诚度以及市场竞争力。

问卷调查是获取用户满意度数据的重要工具, 也是最常见的方法[7]。企业可以通过使用已经检验过的问卷, 多种渠道进行问卷发放, 如在线平台、电话调查、线下活动等, 深入了解用户对产品或服务的看法及感受。问卷内容通常会包括对产品或服务的整体评价、特定功能或服务的满意度、使用频率、问题反馈等方面问题。这些反馈数据不仅有助于企业识别产品或服务中存在的问题和不足, 还能为企业提供产品改进的方向和思路, 同时这些数据还可以作为企业制定市场策略的重要依据, 帮助企业更好地了解市场需求和竞争态势, 提高产品或服务的竞争力。随着市场竞争的加剧, 用户满意度已成为产品核心竞争力的重要组成部分, 因此, 深入研究用户满意度, 提高产品或服务的竞争力十分重要。

场景后问卷[8] (After-Scenario Questionnaire, ASQ)由 Lewis 在 1991 年开发, 主要用于评估用户在完成特定任务或场景后的满意度和感知可用性, 适用于各种用户研究场景, 如软件开发、产品设计、界面测试等, 以了解用户在使用产品或系统过程中的实际体验和感受。其可靠性分析 alpha 系数为 0.95。

用户界面满意度问卷[9] (Questionnaire for User Satisfaction, QUIS)主要用于评估用户界面的满意度, 由马里兰大学的人工智能实验室于 1998 年编制而成, 可靠性分析 alpha 系数为 0.94, 具有较高的通用性, 目前国内没有对本量表进行标准的本土化。QUIS 的应用群体包括: 学生、计算机专业人员、计算机爱好者、新手用户, 通过五个维度对用户界面进行评价: (1) 总体反应: 评估用户对界面的整体感受; (2) 屏幕设计: 包括字体易读性、页面布局、颜色对比度等方面的评价; (3) 术语/系统信息: 评估系统术语是否清晰易懂, 以及系统提供的信息是否充足和准确; (4) 学习: 评估用户学习使用系统的难易程度, 包括初期学习、高级功能学习等; (5) 系统能力: 评估系统响应速度、错误处理、用户控制等方面的性能[9]。问卷设计充分考虑了用户界面的多个关键方面, 测评内容能够全面而深入地反映用户对 Windows 11 系统等现代操作系统的满意度水平, 为提升用户体验提供有力的数据支持。

因此, 随着国内用户体验研究的不断深入和本土化需求的不断增加, CSUQ、ASQ、QUIS 这些国际上流行的经典量表, 适合对 Windows 11 系统进行系统测评, 需要进行中国文化的适应性验证, 才能在国内得到更广泛的应用和推广。此外, 由于 Windows 11 系统的发行时间不长, 国内对 Windows 11 系统的可用性和用户满意度研究也较少, 需要使用验证后的中文版 CSUQ、ASQ、QUIS 三个量表对 Windows 11 系统进行现状调查, 以期探索其用户使用特点。

2. 研究方法

2.1. 研究对象

本研究采取线上与线下相结合的方式发放问卷, 测试数据用于对 CSUQ、ASQ 和 QUIS 量表的适用性检验和现状调查, 分别对应六个不同的样本: (1) 样本一, 共收集到 680 份有效问卷, 男性被试 425 名, 女性被试 255 名, 平均年龄在 19~25 岁之间, 且大部分为本科学历, 占比 76.76%。其中, 所使用的 Windows 11 系统电脑类型为笔记本的被试占比 69.85%, 电脑所属品牌排前三的是联想(36.91%)、华为(19.12%)、戴尔(13.82%)。(2) 样本二, 共回收 715 份有效问卷, 男性被试 451 名, 女性被试 264 名, 平均年龄同样也在 19~25 岁, 且 86.85% 的被试拥有本科及以上学历。(3) 样本三, 共得到 674 份有效问卷, 男性被试 419 名, 女性被试 255 名, 平均年龄在 19~25 岁, 绝大部分为本科学历。(4) 样本四(用于 CSUQ 量表的现状调查), 共得到有效问卷 534 份, 男性被试 279 名, 女性被试 255 名。其中, 学生占比 58.05%, 在职人员 41.95%; 70% 以上的被试年龄为 19~25 岁之间, 且为本科学历; 理工类专业有 298 人, 人文社科类有 236 人。(5) 样本五(用于 ASQ 量表的现状调查), 共得到有效样本 547 份, 男性被试 283 名, 女性被试 264 名。其中, 学生占比 59.60%, 在职人员 40.40%, 年龄为 19~25 岁之间的被试占比 71.85%, 本科学历占比 77.88%。理工类专业有 307 人, 人文社科类有 239 人。(6) 样本六(用于 QUIS 量表的现状调查), 共得到有效样本 432 份, 男性被试 214 名, 女性被试 218 名。其中, 学生占比 58.10%, 平均年龄在 19~25 岁之间, 本科学历占比 77.55%。理工类专业有 221 人, 人文社科类有 211 人。

2.2. 研究工具

2.2.1. CSUQ 量表

计算机系统可用性问卷[5] (CSUQ)是在研究后系统可用性问卷(Post-Study System Usability Questionnaire, PSSUQ)的基础上, 由 Lewis 调整改编而成, 旨在更好地适应非实验室环境下的测试需求。因此, CSUQ 在核心结构上与 PSSUQ 还是保持了高度的一致性, 仅在细节表述上做了些许调整。以第 3 题为例, PSSUQ 表述为: “使用这个系统我能快速完成任务”, CSUQ 表述为: “使用这个系统我能快速完成我的工作”。本研究采用的是 CSUQ (第 3 版), 共 16 个项目, 全面评估用户对计算机系统可用性的看法。参与者被要求从“非常不同意”到“非常同意”(7 点评定量表)上对他们的同意程度进行评分, 这些评估项目最终汇总为四个关键维度: 一个整体和三个分量表分数, 分别代表系统整体可用性评价(题项 1~16)、系统有用性(题项 1~6)、信息质量(题项 7~12)以及界面质量(题项 13~15), 计算规则是通过对各相应题项的反应平均值得出各项分数, 为用户提供了一个全面而细致的系统可用性评估视角。

2.2.2. ASQ 量表

场景后问卷[8] (ASQ)由 PSSUQ 和 CSUQ 的作者发表。ASQ 的开发与 PSSUQ 采用相同的形式, 总共包括三个项目, 分别测量用户在三个方面的满意度: 任务难度、完成效率和支持信息(在线帮助、信息和文档)。其项目采用从 1 (非常不同意)到 7 (非常同意)的七点计分。ASQ 整体分数即是 3 个项目得分的平均分。Lewis 的研究指出 ASQ 分数与 PSSUQ 分数之间存在 $r = 0.8$ 的强相关, 与场景任务的成功率也存在 $r = -0.4$ 显著相关。

2.2.3. QUIS 量表

用户界面满意度问卷[9] (QUIS)是由 UMCP 人机交互实验室(HCIL)的多学科研究团队开发的工具,旨在评估用户对人机界面不同方面的主观满意度。QUIS 包括常用的简版和更为全面的长版,其中 QUIS 第 5 版共 27 个项目,包括五个类别:总体反应(Overall Reaction)、屏幕(Screen)、术语/信息系统(Terminology/System Information)、学习性(Learning)、系统能力(System Capabilities)。所有项目均是采用 9 点的语义差异评级,一般不进行整体的加总,可以通过计算各个分类的平均分和标准差来进行对比。

2.3. 效标选择

系统可用性量表[4] (System Usability Scale, SUS)由 Brooke 在 1986 年开发并于 1996 年正式发表,已成为评估系统或产品感知可用性的主流工具之一。该量表共 10 个条目,采用 Likert 5 点计分法(从“非常不同意”到“非常同意”)。计算 SUS 得分时,首先需要将每个条目的原始分转换为 0 至 4 之间的分值,其中偶数题项(反向条目)通过 5 减去原始分得出,奇数题项(正向条目)是原始分减 1,所有转换分相加后乘以 2.5,得出总分,总分反映系统的总体可用性,分数越高表示产品的可用性被评价得越高。

据 Sauro 和 Lewis 的研究报告显示,约 43%的可用性研究采用了 SUS 量表进行正式或非正式测试[10],在国内,Wang 等人[11]对 SUS 进行了中文版的再翻译与验证,结果显示修订后的中文版在跨文化适用性上优于原始版本。

本研究选用 Wang 等人修订后的中文版 SUS 量表作为效标问卷,以检测 CSUQ、ASQ 和 QUIS 量表的效标关联效度。在三个量表的样本中,该量表的 Cronbach's alpha 系数分别为 0.765、0.746 和 0.743,omega 系数为 0.772、0.744 和 0.748,均大于 0.7,均表现出良好的内部一致性。

2.4. 量表的翻译与回译

本研究对 CUSQ、ASQ、QUIS 三个量表进行了翻译与回译的工作:首先由三位语言学教授将英文量表精准翻译成中文,再由一位带有英语背景的心理教授对比三位专家翻译的中文稿,形成初步的中文量表。为检验翻译的准确性,接着由一位英语专业的教授将初稿进行回译,再由一位心理学教授将原版英文量表和回译后的量表进行对比,对中文版量表进行修改调整。最后为进一步优化中文量表的翻译,邀请了 10 位 Windows 11 系统用户进行量表填写,根据用户反馈的建议再次修改翻译,最终形成中文版的 CSUQ、ASQ、QUIS 量表。

3. 数据分析与结果

3.1. 中文版 CUSQ、ASQ、QUIS 三个量表的适用性检验

3.1.1. CUSQ 量表

(1) 项目分析

使用第一次测试所获得的样本一数据($n = 680$),按照 CSUQ 计算规则计算其整体得分之后,对该量表的所有题项与其整体得分做相关分析,结果表明,所有题项与整体得分的相关在 0.01 水平上达到显著,相关系数在 0.635 至 0.722 之间。然后根据整体得分对样本进行高低排序,按前后各 27%划分为高低分组,并对两组样本进行独立样本 t 检验。结果显示,两组样本在 16 个题项的得分均差异显著($p < 0.001$),这说明中文版 CSUQ 量表的 16 个题项具有较好的区分度。

(2) 探索性因子分析

首先, 对于 CSUQ 量表整体, 我们使用样本一的第一部分数据($n=315$)进行 KMO 及 Bartlett 球形检验, 结果表明, $KMO=0.952$, Bartlett 球形检验结果达到显著性水平($\chi^2=4808.269$, $df=120$, $p<0.001$), 说明施测样本数据适合进行探索性因子分析(Exploratory Factor Analysis, EFA) [12]。据此采用最大似然法进行因素抽取, 并在旋转时采用了最大方差法。为了确定因素的数量, 以特征值大于 1 作为依据, 最终提取出一个因子, 累积解释总变异量的 43.221%, 因子负荷在 0.590~0.705 之间(见表 1)。该结果符合将量表作为整体, 用整体得分评估用户对计算机系统可用性的看法。

Table 1. Analysis results of the overall EFA of CSUQ scale

表 1. CSUQ 量表整体探索性因子分析结果

项目	共同度	因子 负荷	特征值	贡献 率	累计 贡献率
整体	整体 满意度		7.480	43.221	43.221
CSUQ1	0.409	0.639	整体上, 我对使用这个系统的容易程度是满意的。		
CSUQ2	0.476	0.690	使用这个系统很简单。		
CSUQ3	0.404	0.636	使用这个系统我能快速完成任务。		
CSUQ4	0.488	0.699	使用这个系统我感觉舒适。		
CSUQ5	0.469	0.685	学习使用这个系统很容易。		
CSUQ6	0.412	0.641	我相信使用这个系统能提高产出。		
CSUQ7	0.360	0.600	这个系统给出的错误提示可以清晰地告诉我如何解决问题。		
CSUQ8	0.348	0.590	我使用这个系统无论何时出错, 都能轻松快速恢复。		
CSUQ9	0.456	0.675	这个系统提供的信息很清晰(如在线帮助、屏幕信息和其他文档)。		
CSUQ10	0.415	0.644	找到我需要的信息很容易。		
CSUQ11	0.428	0.654	信息能够有效地帮助我完成任务。		
CSUQ12	0.436	0.660	系统屏幕上的信息组织很清晰。		
CSUQ13	0.419	0.647	这个系统的界面让人很愉悦。		
CSUQ14	0.425	0.652	我喜欢使用这个系统的界面。		
CSUQ15	0.476	0.690	这个系统具有我期望的所有功能和能力。		
CSUQ16	0.496	0.705	总体来说, 我对这个系统是满意的。		

其次, 对于 CSUQ 三个分量表, 采用分层面因素分析法分别进行探索性因素分析, 具体步骤同上。结果表明, 系统有用性分量表和信息质量分量 KMO 值大于 0.8, 界面质量分量表 KMO 值大于 0.6, 且三个分量表的 Bartlett 球形检验结果均达到显著性水平(KMO 系统有用性 = 0.826, $\chi^2=1248.979$, $df=15$, $p<0.001$; KMO 信息质量 = 0.849, $\chi^2=1005.940$, $df=15$, $p<0.001$; KMO 界面质量 = 0.647, $\chi^2=309.843$, $df=3$, $p<0.001$), 可勉强进行探索性因素分析[13]。据此, 采用最大似然法进行因素抽取, 并在旋转时采用了最大方差法。为了确定因素的数量, 我们以特征值大于 1 作为依据, 每个分量表最终都只提取出一个因子, 累积解释总变异量分别为 43.422%、39.676%、41.208%, 因子负荷分别在 0.613~0.701、0.592~0.668、0.525~0.729 之间(见表 2)。

Table 2. Exploratory factor analysis results of each subscale of the CSUQ scale
表 2. CSUQ 量表各分量表探索性因子分析结果

项目	共同度	因子 负荷	特征值	贡献率	累计贡 献率
分量表 1	系统 有用性		3.168	43.422%	43.422%
CSUQ1	0.435	0.660	整体上, 我对使用这个系统的容易程度是满意的。		
CSUQ2	0.482	0.695	使用这个系统很简单。		
CSUQ3	0.376	0.613	使用这个系统我能快速完成任务。		
CSUQ4	0.492	0.701	使用这个系统我感觉很舒适。		
CSUQ5	0.395	0.629	学习使用这个系统很容易。		
CSUQ6	0.425	0.652	我相信使用这个系统能提高产出。		
分量表 2	信息 质量		2.982	39.676%	39.676%
CSUQ7	0.398	0.631	这个系统给出的错误提示可以清晰地告诉我如何解决问题。		
CSUQ8	0.377	0.614	我使用这个系统无论何时出错, 都能轻松快速恢复。		
CSUQ9	0.446	0.668	这个系统提供的信息很清晰(如在线帮助、屏幕信息和其他文档)。		
CSUQ10	0.404	0.635	找到我需要的信息很容易。		
CSUQ11	0.350	0.592	信息能够有效地帮助我完成任务。		
CSUQ12	0.406	0.637	系统屏幕上的信息组织很清晰。		
分量表 3	界面 质量		1.806	41.208%	41.208%
CSUQ13	0.428	0.655	这个系统的界面让人很愉悦。		
CSUQ14	0.276	0.525	我喜欢使用这个系统的界面。		
CSUQ15	0.532	0.729	这个系统具有我期望的所有功能和能力。		

(3) 验证性因子分析

依据探索性因子分析结果, 我们使用样本一的第二部分数据($n = 365$)进行验证性因子分析(Confirmatory Factor Analysis, CFA), 以检验模型的拟合程度。结果显示, 界面质量分量表的模型达到饱和状态, 整体量表与系统有用性、信息质量分量表的 χ^2/df 分别为 2.220、5.544、2.465, SRMR 均小于 0.8, CFI、TLI 均大于 0.9 (见表 3)。表明各拟合指标结果均表现理想, CSUQ 量表的各结构模型拟合效果良好。

Table 3. The fitness index of CSUQ scale model
表 3. CSUQ 量表模型拟合指数

量表	χ^2	df	p	CFI	TLI	RMSEA (90% C.I.)	SRMR
整体	230.896	104	<0.001	0.951	0.943	0.058 (0.048, 0.068)	0.036
系统有用性	49.896	9	<0.001	0.940	0.900	0.112 (0.083, 0.143)	0.038
信息质量	22.189	9	<0.001	0.976	0.960	0.063 (0.030, 0.097)	0.029
界面质量	0.000	0.000	<0.001	1.000	1.000	0.000 (0.000, 0.000)	0.000

(4) 信度分析

依托 CSUQ 量表可看作一个整体和三个分量表的组成结构, 本研究使用样本一的第二部分数据($n = 365$)对其进行信度检验, 总量表与系统有用性、信息质量两个分量表的内部一致性系数(α)、 ω 系数(ω)均高于 0.8, 界面质量分量表约为 0.7 左右。其次, 总量表与三个分量表的合成信度(CR)分别为 0.927、0.828、0.802 和 0.694。两周后, 在样本一第二部分数据中随机选择 60 人进行了重测, 结果表明, 总量表重测信度为 0.945, 表现良好(见表 4)。

Table 4. The reliability analysis of CSUQ scale

表 4. CSUQ 量表信度分析

维度	克隆巴赫 Alpha (α)	omega 系数 (ω)	合成信度(CR)	重测信度	项数
量表整体	0.927	0.927	0.927	0.945	16
系统有用性	0.827	0.828	0.828	0.877	6
信息质量	0.802	0.802	0.802	0.873	6
界面质量	0.693	0.695	0.694	0.758	3

(5) 效度分析

选用 SUS 量表作为效标问卷, 测得 CSUQ 量表整体和各分量表得分与该效标的得分呈显著的正相关($r_{\text{整体}} = 0.706$, $r_{\text{系统有用性}} = 0.656$, $r_{\text{信息质量}} = 0.675$, $r_{\text{界面质量}} = 0.667$), 说明该量表效标效度良好。

3.1.2. ASQ 量表

(1) 项目分析

该研究过程同上, 我们使用样本二($n = 715$)进行相关分析, 结果显示, ASQ 量表三个题项与其整体得分之间在 0.01 水平上达到显著, 相关系数分别为 0.740、0.796 和 0.837。对高分组与低分组样本在 3 个题项的得分进行独立样本 t 检验的结果表明, 两组样本在 3 个题项的得分均差异显著($p < 0.001$), 这说明中文版 ASQ 量表所有题项具有较好的区分能力。

(2) 信度分析

ASQ 量表的内部一致性系数(α)为 0.699, ω 系数(ω)和合成 w 信度(CR)均为 0.720。由于本量表由 3 道题项组成, 用于探测整体上完成任务的难易度、完成时间和支持信息的满意度。根据杨强等人单维测验合成信度[14](CR)的方法, 计算出总量表的合成信度(CR)为 0.720。同样在时间间隔两周之后, 我们在样本二中随机抽取 60 人进行了重测, 结果为 0.890, 表明重测信度良好。

(3) 效度分析

同样选择 SUS 量表作为效标, 测得 ASQ 量表得分与该效标的相关系数为 0.725, $p < 0.01$, 说明该量表与 SUS 量表间具有较好的效标关联效度。

3.1.3. QUIS 量表

(1) 描述统计

通过计算 QUIS 量表各个类别的平均分和标准差对比发现, 五个类别的均值在 6.93~6.99 之间, 大于中间值 4.5, 标准差在 1.12~1.23 之间, 表明用户对 Windows 11 计算机系统不同方面的主观满意度均较为理想(见表 5)。

Table 5. The descriptive statistical results of the QUIS scale

表 5. QUIS 量表的描述统计结果

	M	SD	N
总体反应	6.99	1.12	674

续表

屏幕	6.98	1.23	674
术语/信息系统	6.96	1.16	674
学习性	6.93	1.20	674
系统能力	6.97	1.15	674

(2) 信度分析

如表 6 所示, QUIS 量表各分类的内部一致性系数(α)和 omega 系数(ω)均高于 0.8; 合成信度(CR)分别为 0.867、0.819、0.872、0.913、0.851。

两周后, 在样本三中随机抽取 60 人进行重测, 结果显示, 除“系统能力”的重测信度为 0.667 外, 其余分类的重测信度均大于 0.8, 表明重测信度较好。

Table 6. The reliability analysis of QUIS scale
表 6. QUIS 量表信度分析

	克隆巴赫 Alpha (α)	omega 系数(ω)	重测信度	合成信度(CR)	项数
总体反应	0.872	0.873	0.912	0.867	6
屏幕	0.829	0.830	0.899	0.819	4
术语/信息系统	0.867	0.868	0.888	0.872	6
学习性	0.879	0.880	0.876	0.913	6
系统能力	0.849	0.851	0.667	0.851	5

(3) 效度分析

① 结构效度

因子分析是直接探讨可观测变量与潜在变量关系的统计方法, 在结构效度的验证中具有独特的价值[15], 分为探索性因子分析(EFA)和验证性因子分析(CFA)两种。目前, 在量表结构效度的评价中两种结合应用得较多。但如果研究者事先对于量表的内在结构已经有了一定的了解, 也就是说, 已经清楚哪些可观测变量可能被哪一个潜在因子所影响, 只需进一步确定每一个潜在因子对可观测变量的影响程度, 这时可以用 CFA 进行[16]。验证性因子分析可以看成是结构方程模型(Structural Equation Model, SEM)的测量模型[17], 采用 Mplus 软件对 QUIS 量表的 5 个分类进行因子结构效度分析, 实质就是对 SEM 的测量模型进行的可靠性评价, 其参数估计方法以及拟合思想与结构方程模型相同。

在本研究中, 使用样本三($n=674$)对 QUIS 量表的理论模型进行验证, 结果表明, 各条目的标准化因子载荷间于 0.682~0.771 之间。模型的主要拟合指标(见表 7): χ^2/df 为 4.400, 介于 3 和 5 之间, 虽然略大于 3, 但由于样本量较大, 仍然符合拟合指数的基本要求; RMSEA 为 0.071, SRMR 为 0.037, 均小于 0.08, 满足拟合指数的要求; CFI 和 TLI 分别为 0.918 和 0.908, 均大于 0.9。综上结果表明模型拟合效果良好, 该量表的结构效度可以接受。

Table 7. The fitness index of QUIS scale model
表 7. QUIS 量表模型拟合指数

χ^2	df	p	CFI	TLI	RMSEA (90% C.I.)	SRMR
1381.746	314	<0.001	0.918	0.908	0.071 (0.067, 0.075)	0.037

② 聚合效度

由表 8 可知, 该量表中, 除了第 3 题和第 8 题之外, 每个分类下对应的题目因子负荷都大于 0.7, 这表明它们在所属类别具有较高的代表性。此外, 每个类别的平均方差变异(AVE)值都大于 0.5, 且组合信度(CR)值均大于 0.8, 这表明该量表的聚合效度[18]非常理想。

Table 8. Factor loading of QUIS scale

表 8. QUIS 量表因子负荷数表

	路径		Estimate	AVE	CR
QUIS1	<---	总体反应	0.709	0.521	0.867
QUIS2	<---	总体反应	0.739		
QUIS3	<---	总体反应	0.688		
QUIS4	<---	总体反应	0.743		
QUIS5	<---	总体反应	0.728		
QUIS6	<---	总体反应	0.724		
QUIS7	<---	屏幕	0.753	0.532	0.819
QUIS8	<---	屏幕	0.682		
QUIS9	<---	屏幕	0.725		
QUIS10	<---	屏幕	0.754		
QUIS11	<---	术语/信息系统	0.752	0.531	0.872
QUIS12	<---	术语/信息系统	0.725		
QUIS13	<---	术语/信息系统	0.723		
QUIS14	<---	术语/信息系统	0.703		
QUIS15	<---	术语/信息系统	0.761		
QUIS16	<---	术语/信息系统	0.708		
QUIS17	<---	学习性	0.724	0.538	0.913
QUIS18	<---	学习性	0.732		
QUIS19	<---	学习性	0.719		
QUIS20	<---	学习性	0.747		
QUIS21	<---	学习性	0.771		
QUIS22	<---	学习性	0.705		
QUIS23	<---	系统能力	0.743	0.533	0.851
QUIS24	<---	系统能力	0.716		
QUIS25	<---	系统能力	0.735		
QUIS26	<---	系统能力	0.707		
QUIS27	<---	系统能力	0.747		

③ 效标关联效度

与前相同, 选择 SUS 量表作为效标, 测得 QUIS 量表各类别的得分与该效标的相关系数分别为 0.70、0.69、0.67、0.66 和 0.68, $p < 0.01$, 表明其效标效度可接受。

3.2. Windows 11 系统可用性及用户满意度现状调查

3.2.1. 三个量表总体描述性统计结果

如表 9 所示, 在三个现状调查样本中, Windows 11 系统用户所使用的电脑类型 70%以上都为笔记本, 少数人是台式电脑。另外, 联想和华为是他们使用最多的品牌。其中, 联想的小新系列产品型号最受欢迎, 均有 20%以上的用户在使用。

Table 9. Statistics of the computer conditions of Windows 11 system users in the three samples
表 9. 三个样本中 Windows 11 系统用户的电脑情况统计

样本四(CSUQ, $n = 534$)			样本五(ASQ, $n = 547$)			样本六(QUIS, $n = 432$)		
类别	频率	占比	类别	频率	占比	类别	频率	占比
电脑类型								
笔记本	377	70.60%	笔记本	401	73.31%	笔记本	324	75.00%
平板	18	3.37%	平板	125	22.85%	平板	8	1.85%
台式电脑	139	26.03%	台式电脑	20	3.66%	台式电脑	100	23.15%
电脑品牌(前四)								
联想	187	35.02%	联想	192	35.10%	联想	164	37.96%
华为	101	18.91%	华为	99	18.10%	华为	75	17.36%
戴尔	76	14.23%	戴尔	81	14.81%	戴尔	65	15.05%
惠普	65	12.17%	惠普	63	11.52%	惠普	49	11.34%
电脑型号(前四)								
小新系列	112	20.97%	小新系列	121	22.12%	小新系列	98	22.69%
KLVG-32	28	5.24%	KLVG-32	29	5.30%	Inspiron 系列	22	5.09%
Inspiron 系列	26	4.87%	Inspiron 系列	29	5.30%	KLVG-32	19	4.40%
Iecoo AIO 酷 2888	21	3.93%	戴尔 Vestro	20	3.66%	戴尔 Vestro	18	4.17%

此外, 在三个量表的得分分布中, CSUQ 量表整体及三个分量表的得分均值在 5.58~5.70 之间, 大于中间值 4; ASQ 量表的整体得分均值为 5.51, 也大于中间值 4; QUIS 量表五个分类的得分均值在 6.89~6.96 之间, 大于中间值 4.5 (见表 10)。以上结果均表明, 用户对 Windows 11 系统的体验满意度评价较为理想。

Table 10. Descriptive statistics of the three scales
表 10. 三个量表的描述统计

量表	维度	Min	Max	M $\pm$ SD	n
CSUQ 量表	整体	1.94	7.00	5.64 $\pm$ 0.72	534
	系统有用性	2.33	7.00	5.70 $\pm$ 0.74	534
	信息质量	1.67	7.00	5.58 $\pm$ 0.76	534
	界面质量	1.00	7.00	5.59 $\pm$ 0.83	534
ASQ 量表	整体	1.00	7.00	5.51 $\pm$ 0.85	547
QUIS 量表	总体反应	0.67	9.00	6.89 $\pm$ 1.26	432
	屏幕	0.25	9.00	6.96 $\pm$ 1.35	432

续表

QUIS 量表	术语/信息系统	0.83	9.00	6.92 ± 1.27	432
	学习性	1.00	9.00	6.89 ± 1.32	432
	系统能力	0.80	9.00	6.94 ± 1.27	432

3.2.2. 差异分析

(1) CSUQ 量表

① 多因素方差分析

在性别上, CSUQ 量表整体及三个分量表均不存在显著差异。

在职业与专业上, CSUQ 量表整体及三个分量表得分均交互作用显著($F_{\text{整体}(1,526)} = 6.488, p = 0.011, \eta^2 = 0.012$; $F_{\text{系统有用性}(1,526)} = 4.606, p = 0.032, \eta^2 = 0.009$; $F_{\text{信息质量}(1,526)} = 7.093, p = 0.008, \eta^2 = 0.013$; $F_{\text{界面质量}(1,526)} = 6.933, p = 0.009, \eta^2 = 0.013$), 且表现出一致性趋势, 即人文社科类在职人员有更高的满意度(如表 11、表 12 所示)。具体表现是: 人文社科类专业中, 在职人员对 Windows 11 系统的评价得分显著高于学生; 在职人员中, 人文社科类专业用户的得分显著高于理工类专业用户的得分。

Table 11. The results of multivariate analysis of variance of the CSUQ scale

表 11. CSUQ 量表多因素方差分析结果

	主体间效应检验							
	整体		系统有用性		信息质量		界面质量	
	F	η^2	F	η^2	F	η^2	F	η^2
性别	0.169	0.000	0.000	0.000	0.222	0.000	0.009	0.000
职业	9.195**	0.017	9.483**	0.018	8.058**	0.015	6.141*	0.012
专业	3.819	0.007	2.391	0.005	4.375*	0.008	5.955*	0.011
性别*职业	0.228	0.000	0.529	0.001	0.209	0.000	0.005	0.000
性别*专业	0.057	0.000	0.259	0.000	0.368	0.001	0.572	0.001
职业*专业	6.488*	0.012	4.606*	0.009	7.093**	0.013	6.933**	0.013
性别*职业*专业	0.335	0.001	0.310	0.001	0.281	0.001	0.000	0.000

注: *在 0.05 水平上显著; **在 0.01 水平上显著。

Table 12. Simple effect test of profession & major

表 12. 职业和专业简单效应检验

量表维度	变异来源	ss	ms	F	显著性	LSD
整体	专业					
	理工类	0.067	0.067	0.131	0.718	
	人文社科类	7.201	7.201	14.131***	<0.001	在职人员 > 学生
	职业					
	学生	0.099	0.099	0.195	0.659	
	在职人员	4.701	4.701	9.224**	0.003	人文社科类 > 理工类

续表

系统有用性	职业	专业					
		理工类	0.260	0.26	0.485	0.487	
		人文社科类	6.638	6.638	12.396***	<0.001	在职人员 > 学生
	专业	职业					
		学生	0.107	0.107	0.200	0.655	
		在职人员	3.324	3.324	6.207*	0.013	人文社科类 > 理工类
信息质量	职业	专业					
		理工类	0.010	0.010	0.017	0.896	
		人文社科类	7.672	7.672	13.740***	<0.001	在职人员 > 学生
	专业	职业					
		学生	0.101	0.101	0.181	0.671	
		在职人员	5.747	5.747	10.292**	0.001	人文社科类 > 理工类
界面质量	职业	专业					
		理工类	0.009	0.009	0.013	0.908	
		人文社科类	8.074	8.074	11.858***	<0.001	在职人员 > 学生
	专业	职业					
		学生	0.014	0.014	0.021	0.886	
		在职人员	7.978	7.978	11.717***	<0.001	人文社科类 > 理工类

注: *在 0.05 水平上显著; **在 0.01 水平上显著; ***在 0.001 水平上显著。

② 非参数检验

在电脑类型上, 由于不同类型的被试数量分布差异较大, 不适宜直接进行参数检验。因此, 我们采取了分组策略, 根据不同类型将被试划分为三组(平板、笔记本、台式电脑)后, 对其在 CSUQ 整体及三个分量表上的得分情况进行了非参数检验——多组独立样本秩和检验, 以确保分析结果的准确性和可靠性。三组被试在 CSUQ 整体及三个分量表上的得分均存在显著差异($\chi^2_{\text{整体}} = 13.344, p = 0.001$; $\chi^2_{\text{系统有用性}} = 12.363, p = 0.002$; $\chi^2_{\text{信息质量}} = 12.757, p = 0.002$; $\chi^2_{\text{界面质量}} = 9.825, p = 0.007$)。

进一步经过两两比较可知, 在 CSUQ 整体和信息质量上, 平板与笔记本用户之间得分差异显著($\chi^2_{\text{整体}} = -2.649, p_{\text{整体}} = 0.024$; $\chi^2_{\text{信息质量}} = -2.949, p_{\text{信息质量}} = 0.010$), 笔记本用户的得分显著高于平板用户; 平板与台式电脑用户之间得分也差异显著($\chi^2_{\text{整体}} = -3.425, p_{\text{整体}} = 0.002$; $\chi^2_{\text{信息质量}} = -3.504, p_{\text{信息质量}} = 0.001$), 台式电脑用户得分显著高于平板用户。但笔记本与台式电脑用户之间的得分并无显著差异($\chi^2_{\text{整体}} = -2.206, p = 0.082$; $\chi^2_{\text{信息质量}} = -1.676, p_{\text{信息质量}} = 0.281$)。总体而言, 平板用户在整体和信息质量上对 Windows 11 系统评价得分最低。

在系统有用性上, 平板与台式电脑用户得分也存在显著差异($\chi^2_{\text{系统有用性}} = -2.994, p = 0.008$), 台式电脑用户得分显著高于平板用户; 笔记本与台式电脑用户得分存在显著差异($\chi^2_{\text{系统有用性}} = -2.631, p = 0.026$), 台式电脑用户得分显著高于笔记本用户; 但笔记本与平板用户得分差异不显著($\chi^2_{\text{系统有用性}} = -2.026, p = 0.128$)。总的来说, 台式电脑用户在其系统有用性的评价上相对较高。在界面质量上, 只有平板与台式电脑用户之间得分有显著差异($\chi^2_{\text{界面质量}} = -2.986, p = 0.008$), 台式电脑的用户得分显著高于平板用户。

(2) ASQ 量表

① 多因素方差分析

不同专业和职业在 ASQ 量表整体得分上交互作用显著($F_{\text{整体}(1,539)} = 5.943$, $p = 0.015 < 0.05$, $\eta^2 = 0.011$), 整体趋势为文科类在职人员对 Windows 11 系统评分最好(见表 13、表 14)。具体表现为: 在职人员中, 人文社科类专业用户整体得分显著高于理工类专业; 人文社科类专业中, 在职人员整体得分显著高于学生用户。

Table 13. The results of multivariate analysis of variance of the ASQ scale

表 13. ASQ 量表多因素方差分析结果

主体间效应检验		
	ASQ 整体	
	F	η^2
性别	0.615	0.001
专业	3.144	0.006
职业	4.134*	0.008
性别*专业	0.001	0.000
性别*职业	0.019	0.000
专业*职业	5.943*	0.011
性别*专业*职业	0.466	0.001

注: *在 0.05 水平上显著。

Table 14. Simple effect test of major and profession

表 14. 专业和职业的简单效应检验结果

	变异来源	ss	ms	F	显著性	LSD
	职业					
专业	学生	0.191	0.191	0.265	0.607	
	在职人员	5.497	5.497	7.611**	0.006	人文社科类 > 理工类
	专业					
职业	理工类	0.066	0.066	0.092	0.762	
	人文社科类	6.527	6.527	9.038**	0.003	在职人员 > 学生

注: *在 0.05 水平上显著; **在 0.01 水平上显著。

② 非参数检验

考虑到所收集到的不同电脑类型的被试量差异较大, 不适宜直接应用参数检验方法。我们同样采用非参数检验中的多组独立样本秩和检验对使用不同电脑类型的三类人群进行比较。结果表明, 三组被试在 ASQ 量表整体上的得分存在显著差异($\chi^2_{\text{整体}} = 13.621$, $p = 0.001$)。

两两比较后发现, 平板与笔记本用户得分之间有显著差异($\chi^2_{\text{整体}} = -3.156$, $p_{\text{整体}} = 0.005$), 笔记本用户得分显著高于平板用户; 平板与台式电脑用户得分之间也有显著差异($\chi^2_{\text{整体}} = -3.669$, $p_{\text{整体}} = 0.001$), 台式电脑用户得分显著高于平板用户。但笔记本与台式电脑用户之间得分并无显著差异($\chi^2_{\text{整体}} = -1.566$, $p_{\text{整体}} = 0.352 > 0.05$)。

(3) QUIS 量表

① 多因素方差分析

性别在 QUIS 量表五个类别得分中均不存在显著差异; 专业和职业在术语/信息系统下也均不存在显著差异。

但职业与专业在总体反应、屏幕、学习性和系统能力四个类别中的得分均表现出交互效应($F_{\text{总体反应}(1,424)} = 6.697, p = 0.010, \eta^2 = 0.016$; $F_{\text{屏幕}(1,424)} = 4.890, p = 0.028, \eta^2 = 0.011$; $F_{\text{学习性}(1,424)} = 3.952, p = 0.047, \eta^2 = 0.009$; $F_{\text{系统能力}(1,424)} = 5.352, p = 0.021, \eta^2 = 0.012$)。在总体反应和屏幕上, 人文社科类的在职人员对 Windows 11 系统的反应评价最好。

具体来说: 人文社科类用户中, 在职人员的各类别得分显著高于学生用户; 在职人员中, 人文社科类专业用户得分显著高于理工类专业用户。在学习性和系统能力上, 人文社科类在职人员好过学生用户(见表 15、表 16)。

Table 15. The results of multivariate analysis of variance of the QUIS scale
表 15. QUIS 量表多因素方差分析结果

	主体间效应检验									
	总体反应		屏幕		术语/信息系统		学习性		系统能力	
	F	η^2	F	η^2	F	η^2	F	η^2	F	η^2
性别	0.651	0.002	0.987	0.002	2.405	0.006	1.851	0.004	0.684	0.002
职业	3.004	0.007	5.155*	0.012	2.945	0.007	3.782	0.009	1.781	0.004
专业	0.748	0.002	1.356	0.003	1.034	0.002	0.728	0.002	0.193	0.000
性别*职业	0.174	0.000	1.322	0.003	0.175	0.000	0.098	0.000	0.088	0.000
性别*专业	0.188	0.000	0.000	0.000	0.117	0.000	0.233	0.001	0.556	0.001
职业*专业	6.697*	0.016	4.890*	0.011	3.178	0.007	3.952*	0.009	5.352*	0.012
性别*职业*专业	0.016	0.000	0.009	0.000	0.124	0.000	0.009	0.000	0.023	0.000

注: *在 0.05 水平上显著。

Table 16. Simple effect test of profession & major
表 16. 职业和专业的简单效应分析

量表维度	变异来源	ss	ms	F	显著性	LSD
总体反应	专业					
	理工类	0.587	0.587	0.374	0.541	
	人文社科类	14.332	14.332	9.125**	0.003	在职人员 > 学生
	职业					
	学生	2.641	2.641	1.681	0.195	
	在职人员	8.381	8.381	5.336*	0.021	人文社科类 > 理工类
屏幕	专业					
	理工类	0.003	0.003	0.002	0.966	
	人文社科类	17.532	17.532	9.817**	0.002	在职人员 > 学生
	职业					
	学生	1.109	1.109	0.621	0.431	
	在职人员	9.109	9.109	5.100*	0.024	人文社科类 > 理工类

续表

学习性	职业	专业					
		理工类	0.002	0.002	0.001	0.975	
		人文社科类	13.074	13.074	7.559**	0.006	在职人员 > 学生
	专业	职业					
		学生	1.261	1.261	0.729	0.394	
		在职人员	6.25	6.25	3.613	0.058	
系统能力	职业	专业					
		理工类	0.786	0.786	0.49	0.484	
		人文社科类	10.426	10.426	6.504*	0.011	在职人员 > 学生
	专业	职业					
		学生	3.19	3.19	1.99	0.159	
		在职人员	5.435	5.435	3.39	0.066	

② 非参数检验

同前述研究, 我们依然采用多组独立样本秩和检验进行差异分析。结果表明: 三组被试在 QUIS 量表各类别上的得分均存在显著差异($\chi^2_{\text{总体反应}} = 15.954, p < 0.001$; $\chi^2_{\text{屏幕}} = 14.840, p = 0.001$; $\chi^2_{\text{术语/信息系统}} = 12.816, p = 0.002$; $\chi^2_{\text{学习性}} = 10.950, p = 0.004$; $\chi^2_{\text{系统能力}} = 13.264, p = 0.001$)。

进一步两两比较分析得出, 在总体反应上, 平板与台式电脑的得分差异显著($\chi^2_{\text{总体反应}} = -3.096, p = 0.006$), 平板用户的得分显著低于台式电脑用户; 笔记本与台式电脑的得分差异显著($\chi^2_{\text{总体反应}} = -3.188, p = 0.004$), 台式电脑用户的得分显著高于笔记本用户; 平板与笔记本的得分差异不显著($\chi^2_{\text{总体反应}} = -2.160, p = 0.092 > 0.05$)。整体上, 台式电脑用户对 Windows 11 系统在总体反应上的评分较高。

在屏幕上, 平板与笔记本、平板与台式电脑、笔记本与台式电脑三组的得分均存在显著差异($\chi^2_{\text{屏幕}} = -2.562, p = 0.031$; $\chi^2_{\text{屏幕}} = -3.329, p = 0.003$; $\chi^2_{\text{屏幕}} = -2.676, p = 0.022$)。整体评分表现为: 台式电脑 > 笔记本 > 平板, 表明 Windows11 系统的屏幕设计可能用于台式电脑更受用户喜欢。

在术语/信息系统和系统能力上, 平板与笔记本用户的得分存在显著性差异($\chi^2_{\text{术语/信息系统}} = -2.631, p = 0.026$; $\chi^2_{\text{系统能力}} = -2.657, p = 0.024$), 笔记本用户在这两个维度上的得分显著高于平板用户; 平板与台式电脑用户的得分也存在显著性差异($\chi^2_{\text{术语/信息系统}} = -3.255, p = 0.003$; $\chi^2_{\text{系统能力}} = -3.299, p = 0.003$), 台式电脑用户的得分显著高于平板用户; 但笔记本与台式电脑之间并无显著性差异($\chi^2_{\text{术语/信息系统}} = -2.222, p = 0.079$; $\chi^2_{\text{系统能力}} = -2.283, p = 0.067 > 0.05$)。总体而言, 台式电脑用户的得分是最好的。

在学习性上, 只有平板与台式电脑用户之间的得分有显著差异($\chi^2_{\text{学习性}} = -2.892, p = 0.011$), 表现为平板用户的得分显著低于台式电脑用户。

4. 讨论

4.1. 三个量表的中文版本验证

分别对三个量表进行项目分析、探索性因子分析、验证性因子分析、信度检验、效度检验等工作, 各项指标达到理想水平。

关于 CSUQ 量表, 项目分析结果显示, 各题项与整体得分的相关系数均在有效范围以内, 具有较好的区分能力。探索性因子分析(EFA)与验证性因子分析(CFA)验证了模型的拟合效果。各项信效度指标均

显示中文版本量表的可靠性较高。

关于 ASQ 量表, 项目分析显示出区分度。三个题项分别聚焦任务完成的核心维度(难度、效率、支持), 逻辑清晰, 符合用户体验评估的理论框架。各项信度指标也达到标准。

在 QUIS 量表上, 项目分析结果显示题项设计合理, 能够有效区分不同满意度的用户群体。

应用验证性因子分析(CFA)对量表 5 个分类进行因子结构效度分析, 模型拟合指标良好, 说明量表的结构效度符合要求。其他各项信效度指标也达到标准。

4.2. Windows 11 系统可用性及用户满意度现状分析

4.2.1. 总体评价

根据 CSUQ、ASQ 和 QUIS 量表的调查结果显示, CSUQ 量表(7 分制)、ASQ 量表(7 分制)、QUIS 量表(9 分制)得分均显著高于中间值, 说明用户对 Windows 11 系统的可用性和满意度评价整体较高。但在事后的抽查回访中也有用户希望 Windows 系统未来可以增强系统的兼容性, 以及提供更友好的学习资源, 进一步提升用户体验。

4.2.2. 差异分析

(1) 在性别上的差异

本研究结果表明, CSUQ、ASQ、QUIS 量表的测评数据反应出性别差异性均不显著, 男女用户在任务完成和界面满意度上的体验比较一致, 这与人研究存在差异[19]。原因可能在于: 随着信息技术和数字化能力的普及, 不同性别的用户在计算机使用习惯和技能掌握上差异逐渐缩小; 现代人机交互(HCI)设计普遍采用性别中立(Gender-Neutral)原则, 所以, 性别对操作系统体验的影响可能相对被弱化。

(2) 职业与专业之间的交互作用

本研究结果表明, 在 CSUQ、ASQ、QUIS 量表的测评数据中, 职业与专业的交互作用均显著, 原因可能在于: ① 人文社科类在职人员因为其工作内容涉及较多的文档处理、信息检索、多任务协作等不同类型的办公场景。他们可能更倾向于图形化界面(GUI)的直观操作, 而非命令行或复杂设置; 注重界面设计, 需要信息呈现清晰的信息架构和视觉层级[20], 因此在系统有用性和界面质量上的评分显著高于其他类用户。② 理工类用户更加地关注系统的功能性和技术性能, 如多任务处理能力、开发环境兼容性、命令行工具支持等。然而, Windows 11 在开发工具集成、虚拟化支持、高级系统管理功能等方面可能未能完全满足其需求, 导致该群体对系统的可用性和满意度评分较低[19]。此外, 理工类用户对技术问题的容忍度也会影响系统可用性及满意度的评分。③ 在职人员通常需要在特定时间内高效地完成工作任务, 因此操作系统时对其响应速度、任务切换流畅度和稳定性要求更高, 学生用户的使用场景更多样化, 会涉及学习、娱乐、社交等, 对系统的多媒体支持(如游戏性能、影音播放)和个性化定制有更高层次的期待, 这些都会导致不同职业与专业的群体在 Windows 11 系统可用性和满意度评分上存在差异。

(3) 在电脑类型上的差异

研究结果发现, 不同电脑类型(笔记本、平板、台式电脑)的用户在 CSUQ、ASQ、QUIS 量表上的评分上多数存在显著性差异, 台式电脑用户在系统可用性和满意度上的评分较高, 笔记本和平板电脑的用户评分次之, 原因可能在于: 台式电脑普遍配备的更高性能硬件, 更强大的处理器、更大的存储空间, 以及更优质的显示设备, 可同时打开多个浏览器窗口、文档和通讯软件, 还能保持界面响应速度(延迟 < 100 ms), 这些优势为用户提供了更为流畅的操作体验和更高的工作效率[21]。相比之下, 笔记本电脑虽然在便携性上占优势地位, 其轻薄设计和内置电池满足了用户“随时办公”的需求, 但如果长时间高负载运行或执行图形密集型任务时, 可能会受到散热和电池续航的限制, 影响用户体验。而平板电脑, 尽管在触控操作、手写笔输入和移动性方面表现出色, 但其操作屏幕较小, 硬件性能有限, 在处理复杂任

务或多任务管理(如无法同时打开两个完整的 Word 文档, 只能分屏显示半页内容)时显得力不从心, 导致系统可用性和满意度的评分较低。

综上, 不同电脑类型的评分差异本质上是“硬件性能-设计定位-使用场景”三者匹配度的体现: 台式电脑以“性能优先”的设计, 满足了专业办公、高性能计算等需求; 笔记本电脑以“便携与性能平衡”的定位, 适配了移动办公场景; 平板电脑则以“触控与轻量”的优势, 聚焦于娱乐和休闲办公。未来若要缩小不同设备的满意度差距, 需针对各自的瓶颈进行优化——例如, 笔记本电脑可通过“液冷散热”提升高负载性能, 平板电脑可通过“外接扩展坞”扩展屏幕和内存, 从而更好地匹配用户的多元化需求。

5. 结论

(1) 通过对 Windows 11 系统用户进行测评, 验证了 CSUQ、ASQ 和 QUIS 量表在中国具有较高的文化适用性。

(2) 验证后的 CSUQ、ASQ 和 QUIS 量表对 Windows 11 系统用户现状调查表明, 各类用户的总体满意度较好。

(3) 不同性别在评价上无显著差异。台式电脑用户在系统可用性和满意度上的评分较高, 笔记本和平板电脑的用户评分次之。职业与专业存在交互作用, 人文社科类在职人员好过学生用户。

参考文献

- [1] 刘克涛. 基于用户体验度量的人机交互界面可用性研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2018.
- [2] 由芳, 王建民. 可用性测试[M]. 广州: 中山大学出版社, 2017.
- [3] 王宇晖. 系统可用性量表在感知可用性纵向研究中的改进和应用[D]: [博士学位论文]. 武汉: 华中科技大学, 2020.
- [4] Brooke, J. (1996) SUS-A “Quick and Dirty” Usability Scale. In: Jordan, P., Thomas, B. and Weerdmeester, B., Eds., *Usability Evaluation in Industry*, Taylor & Francis, 4-9.
- [5] Lewis, J.R. (2002) Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human-Computer Interaction*, **14**, 463-488. <https://doi.org/10.1080/10447318.2002.9669130>
- [6] 赵宇晴, 阮平南, 刘晓燕, 等. 基于在线评论的用户满意度评价研究[J]. 管理评论, 2020, 32(3): 179-189.
- [7] 黄坚平, 李晋明. 用户满意理念及用户满意度指数在中国的应用[J]. 北京商学院学报, 2000(5): 60-63.
- [8] Lewis, J.R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, **7**, 57-78. <https://doi.org/10.1080/10447319509526110>
- [9] Chin J.P. and Diehl, V.A. (1988) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery.
- [10] Lewis, J.R. (2018) The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, **34**, 577-590. <https://doi.org/10.1080/10447318.2018.1455307>
- [11] Wang, Y., Lei, T. and Liu, X. (2019) Chinese System Usability Scale: Translation, Revision, Psychological Measurement. *International Journal of Human-Computer Interaction*, **36**, 953-963. <https://doi.org/10.1080/10447318.2019.1700644>
- [12] 兰国帅, 钟秋菊, 吕彩杰, 等. 探究社区量表中文版的编制——基于探索性和验证性因素分析[J]. 开放教育研究, 2018, 24(3): 68-76.
- [13] Napitupulu, D., Abdel Kadar, J. and Kartika Jati, R. (2017) Validity Testing of Technology Acceptance Model Based on Factor Analysis Approach. *Indonesian Journal of Electrical Engineering and Computer Science*, **5**, 697-704. <https://doi.org/10.11591/ijeecs.v5.i3.pp697-704>
- [14] 杨强, 叶宝娟, 温忠麟. 用 SPSS 软件计算单维测验的合成信度[J]. 中国临床心理学杂志, 2014, 22(3): 496-498.
- [15] Bollen, K.A. (1989) Structural Equations with Latent Variables. Wiley. <https://doi.org/10.1002/9781118619179>
- [16] Bartholomew, D.J. and Knott, M. (1999) Latent Variable Models and Factor Analysis. John Wiley & Sons, 120-125.
- [17] Byrne, B.M. (1998) Structural Equation Modeling with LIS-REL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming. Lawrence Erlbaum, 827-830.

-
- [18] Fornell, C. and Larcker, D.F. (1981) Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, **18**, 382-388. <https://doi.org/10.2307/3150980>
 - [19] 张颖. 大学生计算机焦虑及其相关因素的研究[D]: [硕士学位论文]. 重庆: 西南大学, 2008.
 - [20] 汪雨薇, 程垫丰. 浅议办公室行政文员的工作特点[J]. 丝路视野, 2018(11): 51.
 - [21] 姚杰, 魏巍. 顾客重购意愿的品牌资产维度研究——以笔记本电脑行业为例[J]. 产业经济研究, 2009(4): 61-65, 94.