

Software Usability Assessment: Comparison of the Questionnaire Versus User Performance

Fatemeh Mazani

Department of Computer Engineering; Faculty of Engineering;
Golestan University; Gorgan, Iran;
Email: fateme.mazani.se@gmail.com

Shiva Vafadar*

Department of Computer Engineering; Faculty of Engineering;
Golestan University; Gorgan, Iran Email: sh.vafadar@gu.ac.ir

Received: 02, Feb. 2018 Accepted: 16, Jun. 2018

Abstract: Usability is an important attribute for software quality that can play a significant role in improving user experience. Various methods have been proposed for evaluating the usability of software systems. Selecting a proper method for collecting information and evaluating the usability of a software application is a challenge due to the variety of methods, and consequently the time and cost required for their implementation.

In this paper, ease of use is considered as a subset of usability and it is defined as the user's effort to perform a task in the software application. We evaluate ease of use based on two methods; questionnaire and users' performance in an empirical study and compare the results of these two methods. In this way, two different versions of an online store application with similar functionalities are produced and the users are asked to perform specific operations in the application and then answer questions regarding the ease of use of various software functionalities. During the experience, the interactions between user and the software are logged and consequently the user's performance is measured based on the successful completion of the operation, the time of the task completion, the number of steps taken by the user, the number of errors, and the time taken to fix the error. These criteria have been selected based on observable criteria and ease of use measurement in relevant research.

The results of this study is based on the data of 48 applicants except one of the functionalities which has been expanded to 70 applicants in order to achieve adequate level of confidence. Our results indicate that the evaluation of the questionnaire does not correlate with the performance of users in the measured criteria. In some functions, the

Iranian Journal of
**Information
Processing and
Management**

Iranian Research Institute
for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 34 | No. 4 | pp. 1849-1878

Summer 2019



* Corresponding Author

performance of the users in two versions of the application do not have a significant difference, but their views are significantly different in the questionnaire. On the other hand, in some other functions, the performance of the users in two versions has a significant difference but the results of the questionnaire does not show a significant difference. Verifying the performance of each user with the results of his/her questionnaire does not indicate correlation between these two methods results. Also, according to the results of this study, the honesty of users in the questionnaire evaluation can be questioned, since 80% of the users who have not performed a functionality rated it higher than good in the questionnaire.

Keywords: Human-Computer Interaction, Usability, Empirical Study, Usage Analysis, Questionnaire

ارزیابی قابلیت استفاده نرم افزار: بررسی مقایسه‌ای پرسشنامه و عملکرد کاربران

فاطمه مازنی

کارشناسی ارشد؛ گروه مهندسی کامپیوتر؛
دانشکده فنی و مهندسی؛ دانشگاه گلستان؛
fateme.mazani.se@gmail.com

شیوا وفادار

استادیار؛ گروه مهندسی کامپیوتر؛
دانشکده فنی و مهندسی؛ دانشگاه گلستان؛
sh.vafadar@gu.ac.ir



دریافت: ۱۳۹۶/۱۱/۱۳ | پذیرش: ۱۳۹۷/۰۳/۲۶ | مقاله برای اصلاح به مدت ۲۱ روز نزد پدیدآوران بوده است.

فصلنامه علمی پژوهشی

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۲۲۵۱-۸۲۳۱

نما به در SCOPUS، ISI، و LISTA

jipm.irandoc.ac.ir

دوره ۳۴ | شماره ۴ | صص ۱۸۴۹-۱۸۷۸

تابستان ۱۳۹۸



چکیده: قابلیت استفاده یکی از معیارهای مهم کیفیت نرم افزار است که می تواند در بهبود تجربیات کاربری نقش بسزایی داشته باشد. روش های مختلفی برای ارزیابی قابلیت استفاده نرم افزار ارائه شده است. با توجه به تنوع روش ها و زمان و هزینه مورد نیاز برای اجرای آن ها، روش مناسب جمع آوری اطلاعات و ارزیابی قابلیت استفاده نرم افزار یکی از چالش های موجود در این حوزه است.

در این پژوهش، سهولت استفاده به عنوان زیرمجموعه ای از قابلیت استفاده که در آن تلاش مورد نیاز کاربر برای انجام یک کار در نرم افزار ارزیابی می شود، بر اساس دو روش نظرسنجی از طریق پرسشنامه و ارزیابی بر اساس عملکرد کاربران در یک آزمایش تجربی ارزیابی و نتایج این دو روش مقایسه شده است. بدین منظور، دو نسخه مختلف از یک نرم افزار فروشگاه اینترنتی با کارکردهای یکسان تولید شده است. از کاربران خواسته شده عملیات مشخصی را در نرم افزار انجام دهند و به سؤالات یک پرسشنامه در ارتباط با سهولت استفاده از کارکردهای مختلف نرم افزار پاسخ دهند. حین انجام کارکردهای مورد نظر و از طریق ثبت وقایع، عملکرد کاربران در استفاده از نرم افزار ثبت و معیارهای انجام موفقیت آمیز عملیات، زمان انجام کار، تعداد گام ها در انجام کار، تعداد خطا و زمان صرف شده در برطرف کردن خطا اندازه گیری شده است. این معیارها بر اساس معیارهای قابل مشاهده و اندازه گیری سهولت استفاده در پژوهش های مرتبط انتخاب شده اند.

نتایج این پژوهش، بر اساس داده های ۴۸ کاربر به دست آمده و در مورد

یک کارکرد تعداد شرکت کنندگان در آزمایش، به علت عدم کفایت داده برای دستیابی به سطح اطمینان مطلوب، به ۷۰ کاربر گسترش یافته است. این نتایج نشان می‌دهد که ارزیابی حاصل از پرسشنامه با عملکرد کاربران در معیارهای مورد سنجش همبستگی ندارد. در برخی کارکردها، عملکرد کاربران در دو نسخه اختلاف معناداری نداشته، اما نظرات آن‌ها در پرسشنامه به صورت معناداری متفاوت بوده است. از سوی دیگر، در برخی دیگر از کارکردها، معیار عملکرد کاربران در دو نسخه دارای اختلاف معناداری است، در صورتی که نتایج نظرسنجی اختلاف قابل توجهی را بیان نمی‌کند. بررسی عملکرد هر کاربر با نتایج پرسشنامه نیز همبستگی میان این اطلاعات را نشان نمی‌دهد. همچنین، طبق بررسی‌های انجام‌شده در این آزمایش، صداقت کاربران در ارزیابی از طریق پرسشنامه می‌تواند مورد تردید باشد. زیرا کاربرانی که موفق به انجام کارکرد نشده‌اند، در بالاتر از ۸۰ درصد از مواقع، سهولت استفاده را بالاتر از خوب ارزیابی کرده‌اند.

کلیدواژه‌ها: تعامل انسان و کامپیوتر، قابلیت استفاده، ارزیابی تجربی، ارزیابی عملکرد، پرسشنامه

۱. مقدمه

امروزه، نرم‌افزار به عنوان زمینه‌ای آشنا و فراگیر در تمامی جنبه‌های زندگی روزمره، از کار و تجارت تا تفریحات گسترش یافته است. فراگیر شدن نرم‌افزار منجر به افزایش تعداد و تنوع کاربران آن شده است. قابلیت استفاده^۱ نرم‌افزار^۱ که در فارسی معادل «کاربردپذیری» و «استفاده‌پذیری» نیز ترجمه شده، یکی از ویژگی‌های کیفیت نرم‌افزار است که در آن استفاده کاربران از نرم‌افزار برای رسیدن به اهداف خود به صورت اثربخش، کارا و با رضایت مد نظر قرار می‌گیرد (ISO/IEC 25010). قابلیت استفاده، یکی از ویژگی‌های مهم در مقبول و فراگیر شدن نرم‌افزار در میان کاربران آن و برتری نسبت به نرم‌افزارهای مشابه است.

در ایران نیز، با توجه به گسترش نرم‌افزارهای بومی در بازار تولید نرم‌افزار و تنوع و رقابت در این بازار، قابلیت استفاده در سال‌های اخیر مورد توجه بیشتری قرار گرفته است. به همین علت، مسئله ارزیابی قابلیت استفاده نرم‌افزارهای بومی تولیدشده به یکی از مسائل مطرح در صنعت نرم‌افزار ایران تبدیل شده است. در این راستا، در سال‌های اخیر در ایران پژوهش‌های مختلفی در زمینه اندازه‌گیری قابلیت استفاده صورت گرفته است (حسینی، شریف و نوکاریزی ۱۳۹۵؛ نوکاریزی، خیرخواه و کیانی ۱۳۹۲؛ ابراهیمی، ایمان‌خان و اسماعیلی ۱۳۹۴؛ تهرانی و جمشیدی ۱۳۹۴؛ مانیان، موسی‌خانی و رحیمیان ۱۳۹۳؛

1. usability

هوشمند مروتی، بخشی و ربیعی ۱۳۹۳؛ پدرام، کریم‌زادگان مقدم و عاشقی ۱۳۹۰؛ زره‌ساز، فتاحی و داورپناه ۱۳۸۵؛ زره‌ساز و فتاحی ۱۳۹۳). در این پژوهش‌ها با استفاده از پرسشنامه، اغلب، نظرات کاربران در ارتباط با سادگی استفاده، رضایت و مفید بودن کارکردها مورد ارزیابی قرار می‌گیرد.

اگرچه در بسیاری از پژوهش‌ها در زمینه ارزیابی قابلیت استفاده نرم‌افزار به کارگیری پرسشنامه روشی متداول است، اما مفید بودن استفاده از پرسشنامه و نتایج آن در برخی موارد مورد سؤال و تردید واقع شده است. «استفن-دیویدویتز» در کتاب جالب خود، با عنوان «همه دروغ می‌گویند» استفاده از پرسشنامه را به چالش می‌کشد: «به‌صورت سنتی، وقتی دانشگاهیان یا افراد شاغل در صنعت می‌خواهند داده جمع‌آوری کنند، به سراغ نظرسنجی می‌روند. بدین ترتیب، داده‌هایی با ساختار مشخص، بر اساس اعداد یا چک‌باکس‌های یک پرسشنامه به دست می‌آید. اما زمان این روش کار سپری شده و روزگار داده‌های تمیز، ساده، ساخت یافته مبتنی بر نظرسنجی به پایان رسیده است. در دوره جدید، ردپاهای درهم‌ریخته که ما در مسیر زندگی از خود به جا می‌گذاریم، منبع اصلی داده‌های ماست. کلمات داده هستند، کلیک‌ها داده هستند، لینک‌ها داده هستند و اشتباهات تاپی داده هستند» (Stephens-Davidowitz 2018). وی همچنین خاطرنشان می‌کند که نتایج بررسی داده‌هایی که از این ردپاها به دست می‌آید، با نتایج نظرسنجی‌ها اختلاف قابل توجهی دارد. زیرا اغلب افراد به دلایل روانشناختی و اجتماعی مختلف ممکن است نظرات واقعی خود را در نظرسنجی‌ها ابراز نکنند (Carter 2007). «جانکو» نیز معتقد است که پرسشنامه عاملی انتزاعی است که ما را از کاربر و وقایعی که در هنگام استفاده رخ می‌دهد، جدا می‌کند و ارتباط نزدیک بین تجربیات کاربری با محصول در حال استفاده را از بین می‌برد (Junco 2013). به‌صورت کلی‌تر، این ایده نیز مطرح شده است که در سایر زمینه‌های تحقیقاتی نیز مقادیری که از طریق اظهارات افراد به دست می‌آید، در بسیاری موارد وقتی با رفتارهای واقعی مقایسه می‌شوند به‌صورتی قابل توجه غیر دقیق هستند (Hald, Overgaard and Grau 2003; Celis-Morales et al. 2012; Otten, Littenberg and Harvey-Berino 2010).

سؤالی که در این پژوهش مورد بررسی قرار گرفته، این است که در صورتی که ما از ردپاهای درهم‌ریخته‌ای که کاربران ما در مسیر استفاده از نرم‌افزار به جا می‌گذارند،

به‌عنوان منبع اصلی ارزیابی استفاده کنیم، آیا نتایج به‌دست آمده با نتایج حاصل از ارزیابی همان نرم‌افزار از طریق نظرسنجی متفاوت خواهد بود؟ این پرسش‌ها برای ما مهم هستند، چون در انتخاب روش مناسب جمع‌آوری اطلاعات و ارزیابی قابلیت استفاده نرم‌افزار نقش کلیدی دارند.

در این پژوهش، قابلیت استفاده نرم‌افزار به‌معنای درجه و اندازه‌ای است که یک محصول یا سیستم نرم‌افزاری در یک حوزه خاص می‌تواند توسط کاربران مشخص برای رسیدن به اهداف به‌صورت اثربخش، کارا و با رضایت مورد استفاده قرار گیرد (ISO/IEC 25010). از طرف دیگر، سهولت استفاده، به رضایت عمومی از نرم‌افزار، به‌ویژه از نظر واسط کاربری بازمی‌گردد (Hornbæk 2006). بنابراین، در این پژوهش سهولت استفاده به‌عنوان زیرمجموعه‌ای از قابلیت استفاده در نظر گرفته می‌شود و در راستای پاسخ به پرسش‌های فوق، دو روش مختلف ارزیابی سهولت استفاده انتخاب و نتایج حاصل از این دو روش با یکدیگر مقایسه شده است: ارزیابی با استفاده از پرسشنامه و ارزیابی بر اساس عملکرد کاربران در استفاده از نرم‌افزار.

انتخاب معیارهای مناسب برای اندازه‌گیری سهولت استفاده در هر یک از این روش‌ها بسیار مهم است. در ارزیابی با استفاده از پرسشنامه معمولاً نظرات کاربر در ارتباط با سهولت کارکرد نرم‌افزار مورد پرسش قرار می‌گیرد (Lewis 2018)، در حالی که در ارزیابی عملکرد کاربران، سهولت استفاده به‌صورت مستقیم قابل اندازه‌گیری نیست و لازم است معیارهای متناظر با آن که قابلیت اندازه‌گیری کمی دارند، انتخاب شوند. بررسی پژوهش‌های مرتبط نشان می‌دهد که معیار سهولت استفاده را می‌توان به‌صورت قابل مشاهده و اندازه‌گیری^۱ مرتبط با زمان انجام کار و نرخ خطا (Davis and Davis 1989) و متناظر با میزان تلاش مورد نیاز کاربر برای انجام کار در نرم‌افزار در نظر گرفت (Hornbæk 2006). بنابراین، در این پژوهش زیرمعیارهایی مانند تعداد گام‌های کمتر برای کامل کردن وظایف، تعداد خطای کمتر، اصلاح خطا به‌صورت سریع و آسان، نیاز به تلاش و زمان کمتر و استفاده موفقیت‌آمیز کارکردهای نرم‌افزار در نظر گرفته می‌شوند و بدین ترتیب، امکان مقایسه نتایج حاصل از دو روش ارزیابی فراهم می‌گردد.

به‌منظور پاسخ به پرسش تحقیق، در این مقاله از ابزار پرسشنامه به‌همراه ارزیابی مکاشفه‌ای بر اساس اطلاعات تعاملات کاربر با نرم‌افزار استفاده شده است. بدین منظور، یک آزمایش تجربی طراحی و اجرا شده است. در این آزمایش از کاربران خواسته شده

1. objective

بعد از انجام مجموعه عملیات مشخص در دو نسخه از یک نرم‌افزار، نظرات خود دربارهٔ سهولت استفاده از نرم‌افزار را در پاسخ به سؤالات پرسشنامه اعلام کنند. تجربیات کاربران بر اساس تعاملات آن‌ها با نرم‌افزار در حین کار با هر یک از نرم‌افزارها بر اساس کلیک‌ها، انتخاب لینک‌های مختلف، و جست‌وجوهای انجام‌شده به‌صورت خودکار ثبت شده است. بدین ترتیب، معیارهای انجام موفق، زمان انجام، تعداد گام‌ها، تعداد خطا و زمان اصلاح خطا اندازه‌گیری شده است. بدین ترتیب، سهولت استفاده از نرم‌افزار بر اساس ردپاهای کاربر در مسیر استفاده از نرم‌افزار، بر اساس معیارهای تعریف‌شده به‌صورت کمی اندازه‌گیری و در پایان، نتایج این دو ارزیابی با یکدیگر مقایسه می‌شود. نتایج این تحقیق، از دو دیدگاه کاربردی و پژوهشی قابل توجه است. از دیدگاه کاربردی، این پژوهش می‌تواند مورد استفادهٔ تولیدکنندگان نرم‌افزار در ایران باشد. چنانچه در بخش ۲ شرح داده خواهد شد، ارزیابی‌های قابلیت استفاده در ایران بیشتر بر اساس پرسشنامه انجام می‌گیرد. ارائهٔ جزییات حاصل از دو روش مختلف ارزیابی قابلیت استفاده می‌تواند در شناخت فرهنگ کاربران و در نتیجه، انتخاب روش مناسب ارزیابی مفید باشد. از طرف دیگر، این پژوهش از جنبهٔ مقایسهٔ قابلیت استفاده نیز حائز اهمیت است، زیرا در آن دو روش مختلف ارزیابی سهولت استفادهٔ نرم‌افزار با یکدیگر مقایسه می‌شود. نتایج این مقایسه نشان می‌دهد که سهولت استفاده از دیدگاه کاربر که در پرسشنامه اظهار می‌کند، لزوماً با معیارهای عملکردی او مانند استفادهٔ موفقیت‌آمیز کارکردهای نرم‌افزار، نیاز به تلاش و زمان کمتر، تعداد گام‌های کمتر برای کامل کردن وظایف، تعداد خطای کمتر و اصلاح خطا به‌صورت سریع و آسان همبستگی ندارد. بدین ترتیب، یا باید معیارهای دیگری برای سهولت استفاده تعریف کرد یا نتایج حاصل از پرسشنامه را مورد تردید قرار داد.

در ادامهٔ این مقاله، پژوهش‌های مرتبط با ارزیابی‌های قابلیت استفاده در بخش ۲ بیان می‌شود. فرایند پژوهش در این تحقیق، شامل مجموعهٔ فعالیت‌های انجام‌شده، معیارهای ارزیابی، آزمایش تجربی و روش‌های تحلیل و بررسی کیفیت داده در بخش ۳ تشریح می‌گردد. در بخش ۴، بررسی مقایسه‌ای همبستگی معیارهای مختلف عملکرد و پرسشنامه در کارکردهای مختلف ارائه می‌شود. در بخش ۵، همبستگی عملکرد هر کاربر با پاسخ‌های او در پرسشنامه مورد بررسی قرار می‌گیرد. در بخش ۶، بحث و تفسیر نتایج بیان می‌شود و در پایان، جمع‌بندی و پیشنهادات برای کارهای آینده ارائه می‌گردد.

۲. پیشینه پژوهش

پژوهش‌های مرتبط با این تحقیق را می‌توان در سه دسته مختلف دسته‌بندی کرد: (۱) ارزیابی قابلیت استفاده با پرسشنامه، (۲) تحقیقات انجام‌شده در ایران در زمینه اندازه‌گیری قابلیت استفاده نرم‌افزار، و (۳) بررسی همبستگی معیارهای مختلف قابلیت استفاده. در ادامه، هر یک از این پژوهش‌ها شرح داده می‌شود.

۲-۱. ارزیابی قابلیت استفاده با پرسشنامه

یکی از روش‌های متداول ارزیابی، استفاده از پرسشنامه آنلاین است. به‌عنوان نمونه، در یک ارزیابی که فقط با استفاده از پرسشنامه انجام شده، بر اساس نظرات ۴۰۰ کاربر، ۳۶ حوزه وب که هر کدام بیشتر از ۳۰۰ نرم‌افزار تحت وب را شامل می‌شد، ارزیابی شده است. در این ارزیابی تأثیر فاکتورهای وضوح اهداف، سرعت دانلود، بازخورد، قابلیت اعتماد، استفاده از رنگ و گرافیک بر قابلیت استفاده با مقادیر لیکرت ۵ امتیازی سنجیده شده است (Nathan and Yeow 2011). در پژوهشی دیگر که از طریق ایمیل کردن پرسشنامه به ۱۰۲ کاربر انجام شده، تأثیر فاکتورهای انتظارات کاربران، گزارش خطاهای قابلیت استفاده، یادگیری، دستورالعمل‌ها و راهنماهای نرم‌افزار بر قابلیت استفاده نرم‌افزارهای متن‌باز مورد بررسی قرار گرفته است (Raza, Capreta & Ahmed 2012). در این پژوهش، فرمولی برای محاسبه قابلیت استفاده بر اساس این فاکتورها ارائه شده است. در پژوهشی دیگر که با استفاده از پرسشنامه انجام شده، اثر زمان پاسخ بر تجربه کاربری در یک ارزیابی تجربی مورد سنجش قرار گرفته است (Rangardt and Czaja 2017)؛ به این صورت که از دو نسخه یکسان از یک نرم‌افزار تحت وب برای کاربران یادگست استفاده شده است، با این تفاوت که زمان پاسخ در یک نسخه، یک ثانیه بیشتر از نسخه دیگر است. در پایان پرسشنامه‌ای برای جمع‌آوری اطلاعات تجربه کاربری به شرکت‌کنندگان داده شده است. نتایج آزمون تی‌تست در این تحقیق نشان داده است که زمان پاسخ بر هیچ‌کدام از جنبه‌های مورد بررسی تجربه کاربری اثرگذار نبوده است. همچنین، با استفاده از آزمون‌های تعقیبی بر اساس سن یا تجربه کار با یادگست نیز تفاوت معناداری بین دو نسخه مشاهده نشد.

در پژوهشی دیگر، تأثیر چهار عامل قابلیت استفاده، سادگی استفاده، سرگرمی و جامعیت نرم‌افزارهای تحت وب برای ارائه خدمات هتل بر اعتماد اینترنتی و رزرو اینترنتی

هتل مورد بررسی قرار گرفته است (Li et al. 2017). برای ارزیابی این معیارها، ۲۹۸ کاربر پرسشنامه دریافت کردند. این پژوهش از روش "SEM" برای ارزیابی مدل تئوری فرضیات خود استفاده کرده است. نتایج این ارزیابی نشان داده است که قابلیت استفاده، سرگرمی و جامعیت نرم‌افزار تحت وب بر اعتماد اینترنتی اثر مثبت داشته و تأثیر جامعیت بیشتر از سایر معیارها بوده است.

یکی از ارزیابی‌های کمی انجام شده، اندازه‌گیری اثرات اصلاح سیستم راهبری برای بهبود قابلیت استفاده در برنامه‌های کاربردی وب است که به ارزیابی معیارهای وظایف انجام شده صحیح، زمان انجام وظایف در ارتباط با معیار اثربخشی، و سهولت استفاده می‌پردازد (Grigera et al. 2016). هدف از اعمال اصلاحات، حل مشکلات مربوط به پیمایش (سیستم راهبری) است. نحوه اعمال و اصلاح این مشکلات در پژوهش Distant (2014) et al. ارائه شده است که در آن سهولت استفاده به‌عنوان یک معیار مؤثر بر قابلیت استفاده در نظر گرفته شده و به کمک پرسشنامه ارزیابی شده است. همچنین، اثر اعمال اصلاحات مذکور بر بهبود سهولت استفاده نرم‌افزار مورد ارزیابی قرار گرفته است.

در یک پژوهش اثرات اعمال دستورالعمل‌های راهنما برای طراحی فرم، مورد ارزیابی قرار گرفته است. بر اساس نتایج حاصل از بررسی راهنماها توسط افراد خبره، سه فرم طبقه‌بندی شده از بد تا خوب انتخاب شده و سپس، ۶۵ کاربر در اجرای ارزیابی تجربی شرکت کرده‌اند. در این پژوهش معیارهای تعداد ثبت فرم، تعداد و نوع خطاها، زمان کامل کردن فرم و معیارهای پایش حرکت چشم و رضایت کاربر با استفاده از پرسشنامه ثبت شده‌اند (Seckler et al. 2014). نتایج نشان می‌دهد که اعمال این راهنماها سبب بهبود در معیارهای قابلیت استفاده شده است. همچنین، مقایسه نظرات افراد خبره و کاربران نشان داد که مشکلاتی که بیشتر توسط کاربران بیان شده بود، از نظر افراد خبره در اهمیت پایینی قرار داشته است. همچنین، در پژوهش Bargas-Avila et al. (2011) به بررسی مقایسه استفاده از مثال و توضیحات در فرم آنلاین پرداخته شده و معیارهای تعداد خطا تا زمان اولین ارسال، تعداد کل خطاها و تعداد ارسال‌ها تا ثبت درست فرم ذخیره می‌شوند. تفاوت بین استفاده از رنگ در برابر ستاره برای مشخص شدن فیلدهای الزامی نیز در پژوهش Pauwels et al. (2009) مورد بررسی بوده است. معیارهای زمان کامل کردن فرم، نرخ خطا و رضایت کاربر از طریق پرسشنامه ارزیابی شده است.

۲-۲. ارزیابی قابلیت استفاده در ایران

کارهای انجام شده در پژوهش‌های داخلی را می‌توان بر مبنای دو روش اصلی استفاده از پرسشنامه یا روش ارزیابی مکاشفه‌ای قرار داد. در بسیاری از تحقیقات مبتنی بر پرسشنامه، قابلیت استفاده به‌عنوان پارامتری از کیفیت خدمات و سرویس‌های رابط کاربر مشخص شده و اثر آن بر رضایت‌مندی و وفاداری مشتریان یا کاربران وبسایت‌ها مورد بررسی قرار گرفته است (تهرانی و جمشیدی ۱۳۹۴؛ ابراهیمی، ایمان‌خان و اسماعیلی ۱۳۹۴؛ بابایی، صالح اصفهانی و نورانی ۱۳۹۲؛ فتاحی، دینانی و حسینی ۱۳۹۲؛ پدرام، کریم‌زادگان مقدم و عاشقی ۱۳۹۰؛ کریم‌زادگان، خداپرست و وحدت ۱۳۹۰؛ زره‌ساز، فتاحی و داورپناه ۱۳۸۵). در برخی مقالات، قابلیت استفاده به‌عنوان معیار اصلی ارزیابی توسط فاکتورهای مکاشفه‌ای چون سازگاری، راهنما، راهبری، اصلاح خطا، تعامل، قابلیت یادگیری، جست‌وجو، نمایش واژگان، زیبایی‌شناختی و غیره مورد بازرسی و رتبه‌بندی قرار گرفته (احمدیان، صالحی، عابدین‌زاده و خطیبی ۱۳۹۶؛ نارمنجی ۱۳۹۶؛ فرقانی و رستگار ۱۳۹۵؛ حسینی، شریف و نوکاریزی ۱۳۹۵؛ هوشمند مروستی، بخشی و ربیعی ۱۳۹۳؛ آقارضایی و همکاران ۱۳۹۲؛ خواجویی، عزیززی و آتشی ۱۳۹۲؛ آخشیک و رجبعلی‌بگلو ۱۳۹۲؛ و حسن‌زاده و اسکندری ۱۳۹۰). در برخی دیگر از پژوهش‌های مورد بررسی، مانند پژوهش‌های «شکاری و افشاری» (۱۳۹۳)، «نوکاریزی، خیرخواه، و کیانی» (۱۳۹۲) و «زره‌ساز و فتاحی» (۱۳۹۳) در ارزیابی با استفاده از پرسشنامه به‌طور صریح، معیارهای قابلیت استفاده مانند بهره‌وری، اثربخشی و رضایت‌مندی مورد سنجش قرار گرفته است. در این پژوهش‌ها، بیشتر نرم‌افزارهای مورد بررسی در حوزه وبسایت‌های کتابخانه‌ها، دانشگاه‌ها و خدمات مربوط به سلامت بوده است.

۲-۳. بررسی همبستگی معیارهای قابلیت استفاده

پژوهش‌های مختلفی به بررسی همبستگی معیارهای قابلیت استفاده پرداخته‌اند. به‌عنوان نمونه، با استفاده از تحقیقات مروری بر مقالات انجام شده در زمینه قابلیت استفاده، همبستگی بین سه معیار اثربخشی، بهره‌وری و رضایت‌مندی مورد بررسی قرار گرفته است (Frøkjær, Hertzum & Hornbæk 2000). نتایج نشان داده است که هیچ همبستگی بین سه معیار مورد بررسی وجود ندارد و هر کدام از این معیارها به‌صورت جداگانه باید محاسبه شود. در این پژوهش، به‌علت عدم دسترسی به داده‌های خام، پژوهش‌های مورد بررسی

فقط بر اساس خلاصه نتایج آماری انجام شده است؛ در صورتی که ارزیابی مجدد بر اساس داده‌های خام پژوهش‌های مورد استناد، همبستگی ضعیفی را بین معیارها نشان داده است (Hornbæk and Law 2007). اشکال این پژوهش در آن است که به علت تنوع در پژوهش‌های مختلفی که داده‌ها در آن‌ها جمع‌آوری شده، بین معیارهای اندازه‌گیری شده در پژوهش‌های مختلف رابطه و پیوستگی وجود ندارد. محدود کردن ارزیابی‌های این پژوهش به داده‌های مرتبط (که فقط شامل ۱۱ درصد داده‌های اصلی است) حاکی از فقدان همبستگی میان بهره‌وری و رضایتمندی کاربران به عنوان زیرمعیارهای قابلیت استفاده است.

۳. فرایند پژوهش

این پژوهش از نظر نوع، کاربردی و از نظر راهکار اجرایی، آزمایش تجربی کنترل شده است. پرسش تحقیق که در این پژوهش مطرح شده این است که آیا ارزیابی سهولت استفاده نرم‌افزار بر اساس پرسشنامه و بر اساس عملکرد کاربران دارای نتایج یکسانی است؟ در این راستا، دو فرض وجود دارد:

- ◇ فرض صفر: میان دو روش اختلاف معناداری وجود ندارد؛
- ◇ فرض خلاف: میان دو روش اختلاف معناداری وجود دارد.

در این بخش، فعالیت‌های انجام‌شده به‌منظور پاسخ به پرسش پژوهش تشریح می‌گردد.

۳-۱. معیارهای ارزیابی

نخستین گام در فرایند پژوهش تعیین معیارهای ارزیابی است. در این پژوهش، تمرکز بر سهولت استفاده به‌عنوان یکی از ویژگی‌های مهم قابلیت استفاده است (ISO/IEC 25010). در روش ارزیابی از طریق پرسشنامه میزان سهولت استفاده از کاربران پرسیده شده است. این روش طراحی پرسشنامه، بر اساس متداول‌ترین پرسشنامه‌های موجود در ارزیابی قابلیت استفاده صورت گرفته است که در آن نظرات کاربران در ارتباط با سهولت استفاده از سیستم به‌صورت مستقیم پرسیده می‌شود (Lewis 2018). بر این اساس، در این پژوهش میزان سهولت استفاده هر کارکرد مورد ارزیابی از کاربر پرسیده شده است. سهولت استفاده اغلب یک مفهوم شهودی است که به‌صورت مستقیم قابل

اندازه‌گیری نیست. در ارزیابی عملکرد کاربران لازم است معیارهای قابل اندازه‌گیری داشته باشیم. بنابراین، لازم است نگراشت میان سهولت استفاده بر اساس معیارهای قابل اندازه‌گیری بر اساس عملکرد کاربران انجام شود. در فضای تحقیق، سهولت استفاده معمولاً با معیارهایی مانند استفاده موفقیت آمیز از کارکردهای نرم‌افزار، نیاز به تلاش و زمان کمتر برای اجرای یک کارکرد، تعداد گام‌های کمتر برای کامل کردن فعالیت‌های یک کارکرد، تعداد خطای کمتر و اصلاح خطا به صورت سریع و آسان تعریف می‌شود (Hornbæk 2006 و Davis and Davis 1989). بنابراین، در این پژوهش معیارهایی برای اندازه‌گیری عملکرد کاربران در نظر گرفته شده است. تعریف این معیارها و روش محاسبه آن‌ها در جدول ۱، نشان داده شده است. اطلاعات موردنیاز برای محاسبه این معیارها به صورت خودکار در هنگام استفاده کاربران از نرم‌افزار جمع‌آوری می‌گردد.

جدول ۱. معیارهای سهولت استفاده در ارزیابی عملکرد کاربران
(Saleh, Isami and Fabil 2015; Hornbæk 2006; ISO/IEC 25010)

شماره	معیار قابلیت استفاده	روش اندازه‌گیری
۱	انجام موفق	تعداد کاربرانی که عملیات مورد نظر را با موفقیت تا آخرین مرحله انجام داده‌اند.
۲	زمان انجام	$T = T_b - T_a$ $T =$ زمان اتمام اجرای کارکرد $T_a =$ زمان شروع اجرای کارکرد $T_b =$ زمان تکمیل کارکرد مورد نظر
۳	تعداد گام‌ها	تعداد صفحات، کلیک بر دکمه‌های صفحه و تعداد تلاش‌های کاربران برای اجرای کارکرد
۴	تعداد خطا	تعداد دفعات نمایش پیام خطا برای کاربران در هنگام اجرای کارکرد
۵	زمان اصلاح خطاها	$T = \sum_{i=1}^n T_{b_i} - T_a$ $T_a =$ زمان مشاهده پیام خطای I-ام $T_b =$ زمان اصلاح خطای I-ام

۲-۳. طراحی و اجرای آزمایش تجربی

در راستای پاسخ به پرسش‌های پژوهش یک ارزیابی تجربی به روش آزمایش کنترل‌شده طراحی و اجرا شده است. بدین منظور، یک نرم‌افزار عمومی و غیرتخصصی انتخاب شده است تا اثرات دانش پایه و تخصص مورد نیاز برای کار با آن، نتایج آزمایش را تحت تأثیر قرار ندهد. در این راستا، یک نرم‌افزار فروشگاه اینترنتی انتخاب شده است. همچنین، برای این که اثرات یادگیری و کارکرد قبلی شرکت‌کنندگان در آزمایش و

علاقه یا عدم علاقه آنها به نرم افزارهای متداول بر نتایج نظرسنجی اثر گذار نباشد، از نرم افزارهای موجود فروشگاه اینترنتی استفاده نشده است و نرم افزار جدیدی برای این آزمایش طراحی و پیاده سازی شده است.

برای این که بتوان مقایسه میان پرسشنامه و عملکرد کاربران را از جنبه‌های مختلف انجام داد، آزمایش بر اساس دو نسخه مختلف از یک نرم افزار طراحی شده است. دو نسخه نرم افزار از جنبه‌های مختلف مانند کارکردها، قابلیت‌ها، منوهای برنامه، مسیر دسترسی، رنگ بندی و ... کاملاً شبیه یکدیگر هستند. تنها تفاوت این دو نسخه در نحوه مدیریت خطا در آنهاست. هر کاربر فقط از یک نسخه از نرم افزار استفاده می کند و نظرات خود را در ارتباط با همان نسخه در پاسخ به پرسشنامه ارائه می دهد. بدین ترتیب، در آزمایش طراحی شده امکان ارزیابی همبستگی نظرات هر کاربر در پرسشنامه و معیارهای عملکرد او در تعامل با همان نسخه از نرم افزار وجود دارد. علاوه بر آن، در آزمایش طراحی شده می توان نتایج نظرسنجی ها برای دو نسخه را با هم مقایسه کرد، نتایج عملکرد کاربران در دو نسخه را با هم مقایسه کرد، و بررسی کرد که آیا تفاوتی که در نظرسنجی مشاهده می شود با تفاوت در عملکرد کاربران در تعامل با نرم افزار همبستگی دارد یا خیر.

بدین منظور، در نسخه ۱ نرم افزار، کارکردهای متداول نرم افزارهای فروشگاه اینترنتی از قبیل جست و جوی کالا، افزودن کالا به سبد خرید، حذف کالا از سبد خرید و ثبت نام خریدار و ... در نظر گرفته شده است. نسخه ۲ نرم افزار نیز دقیقاً دارای همین کارکردهاست، با این تفاوت که در آن روش های بهبود قابلیت استفاده از قبیل الگوهای مدیریت خطا اعمال شده است. الگوهای مدیریت خطا، راهکارهای شناخته شده طراحی واسط کاربر برای کاهش تعداد خطاهایی است که کاربران در تعامل با نرم افزار با آن برخورد می کنند (Rodríguez, Acuña & Juristoa 2015). الگوهای اعمال شده عبارت اند از: ۱) الگوی تکمیل خودکار در کارکرد جست و جوی کالا، ۲) الگوی خنثی سازی در کارکرد بازگرداندن کالا به سبد خرید، ۳) الگوی نگهداری تاریخچه در کارکرد خرید کالای مشاهده شده، ۴) الگوی عملیات گروهی در کارکرد حذف کالاها از سبد خرید، و ۵) الگوی اعتبارسنجی داده در کارکرد تکمیل اطلاعات.

در هر نسخه از نرم افزار، کدهای لازم برای ثبت خودکار تعاملات کاربر با نرم افزار اضافه شده است. زمان دقیق هر فعالیت کاربر به همراه روش انجام آن و پاسخ سیستم در پایگاه داده ثبت شده است. مقادیر معیارهای ارزیابی عملکرد کاربران که در بخش قبل

ذکر شد، بر اساس اطلاعات ثبت شده در پایگاه داده برای هر کاربر استخراج شده است. البته، لازم به ذکر است که دو معیار تعداد خطا و زمان اصلاح خطا فقط در کارکرد تکمیل اطلاعات اندازه گیری شده است. به این علت که در سایر کارکردها، کاربران در صورت انجام اشتباه، پیغام خطا دریافت نمی کنند و به کار خود ادامه می دهند. معیار تعداد گام، به این علت که شامل گام های درست و نادرست کاربر است، دربرگیرنده تعداد اشتباهات کاربر در سایر کارکردهاست.

در اجرای آزمایش از کاربران خواسته شده که عملیات مشخص را انجام دهند. عملیات مورد انتظار در دو نسخه از نرم افزار کاملاً یکسان بوده است. اما کاربران در نحوه اجرای عملیات آزاد بوده و می توانستند از هر مسیر یا قابلیت نرم افزار برای انجام عملیات مورد نظر استفاده کنند. در پایان استفاده از نرم افزار از هر یک از شرکت کنندگان خواسته شده که نظرات خود را با تکمیل پرسشنامه اعلام نمایند. پرسشنامه، شامل سؤالاتی درباره سهولت استفاده هر یک از کارکردهای مورد ارزیابی است. پاسخ ها در قالب مقیاس «لیکرت» ۵ امتیازی (شامل کاملاً مخالفم، مخالفم، نظری ندارم، موافقم و کاملاً موافقم) طراحی شده که به ترتیب، امتیازی برابر عدد ۱ تا ۵ برای آن ها در نظر گرفته شده است. ساختار و سؤالات پرسشنامه نیز در هر دو نسخه یکسان بوده است. پرسشنامه در جدول ۲، نشان داده شده است.

جدول ۲. پرسشنامه ارزیابی سهولت استفاده در کارکردهای مختلف

شماره	لطفاً به سؤالات مطرح شده درباره نرم افزار پاسخ دهید.	کاملاً موافقم	موافقم ندارم	مخالفم	کاملاً مخالفم
۱	فیلد جست و جو در پیدا کردن آسان تر کالاهای به من کمک کرده است.				
۲	به راحتی توانستم کالایی را که به آن امتیاز داده بودم، به سبد خرید خود اضافه کنم.				
۳	پس از حذف کالا توانستم به راحتی آن را به سبد خرید خود برگردانم.				
۴	حذف کالاهای به آسانی انجام شده است.				
۵	نرم افزار توانست در پر کردن راحت تر فرم به من کمک کند.				
۶	پیغام های خطا به وضوح توانست در تصحیح جواب های نادرست فرم به من کمک کند.				
۷	مکان قرار گیری پیغام های خطا مناسب است.				

در اجرای آزمایش، ۴۸ کاربر (۲۹ زن و ۱۹ مرد) شرکت کرده‌اند. این تعداد، برای کارکرد تکمیل اطلاعات به ۳۵ نفر در هر نسخه از نرم‌افزار افزایش یافته است تا از کفایت داده برای نتیجه‌گیری از آزمون‌های آماری اطمینان حاصل گردد. (از این تعداد ۴۴ نفر از شرکت‌کنندگان زن و ۲۶ نفر مرد بوده‌اند). شرکت‌کنندگان همگی از دانشجویان مقطع کارشناسی و کارشناسی ارشد از رشته‌های مهندسی کامپیوتر و مهندسی برق «دانشگاه گلستان» بوده‌اند و بازه سنی آن‌ها بین ۱۸ تا ۳۵ سال بوده است. این افراد به صورت تصادفی به دو گروه مساوی به تعداد ۲۴ نفر تقسیم شده‌اند و هر گروه فقط از یک نسخه نرم‌افزار استفاده کرده است..

۳-۳. ارزیابی کفایت داده و تحلیل نتایج

یکی از چالش‌های ارزیابی تجربی، تعداد نمونه کافی در ارزیابی است که بتوان به نتایج حاصل از آزمون‌های آماری اطمینان نمود. به همین دلیل، در این پژوهش کفایت داده‌های آزمایش برای تأیید یا رد فرضیات آزمایش مورد بررسی قرار می‌گیرد، زیرا این امکان وجود دارد که نبود اختلاف معنادار میان دو نسخه به علت کوچک بودن اندازه نمونه باشد. در نتیجه، علاوه بر سطح معناداری آزمون‌های آماری باید ضریب تأثیر نیز در تحلیل نتایج لحاظ شود. در این صورت، با محاسبه ضریب تأثیر می‌توان تعداد نمونه کافی برای ارزیابی را به دست آورد.

در این پژوهش، از روش نمونه‌گیری «کوهن»^۱ و ابزار "G*Power" برای بررسی کفایت داده و محاسبه تعداد نمونه مورد نیاز استفاده شده است (Faul et al. 2007). بررسی و تحلیل نتایج اولیه آزمایش نشان می‌دهد که برای کلیه کارکردهای مورد ارزیابی، به جز تکمیل اطلاعات، تعداد نمونه مورد استفاده برای رد یا تأیید فرضیات کافی است. اما در تکمیل اطلاعات، تعداد نمونه کافی برای توان آماری ۰/۸، احتمال وقوع خطای نوع اول ۰/۰۵ و ضریب تأثیر ۰/۷، تعداد شرکت‌کنندگان در آزمایش باید برای هر نسخه از نرم‌افزار ۳۵ نفر باشد. در نتیجه، ارزیابی برای بار دوم برای این کارکرد تکرار شده تا تعداد شرکت‌کنندگان به ۷۰ نفر برسد. در سایر کارکردها بر اساس روش نمونه‌گیری «کوهن»، تعداد ۴۸ نفر کفایت می‌کند.

1. Cohen

پس از اطمینان از کفایت داده‌های آزمایش، تحلیل نتایج به روش‌های مختلفی انجام شده و در این کار از آزمون‌های آماری متنوع استفاده شده است. یکی از روش‌های تحلیل، آزمون بررسی میانگین است. در این آزمون بررسی می‌شود که آیا به صورت میانگین بین دو نمونه مورد بررسی اختلاف معناداری وجود دارد یا خیر. برای پاسخ به این پرسش تحقیق، از آزمون «مان-ویتنی» استفاده شده است.

در تحلیل نتایج پرسشنامه، برای پاسخ‌های پرسشنامه اعداد ۱ تا ۵ به‌ازای هر پرسش (عدد ۱ برای کاملاً مخالفم تا ۵ برای کاملاً موافقم) در نظر گرفته شده است. در این آزمایش داده‌های پرسشنامه از نوع متغیر رتبه‌ای است. بنابراین، از آزمون ناپارامتری «مان-ویتنی» در تعیین سطح معناداری اختلاف میان دو نسخه نرم‌افزار استفاده شده است.

در ارزیابی معیارهای مرتبط با عملکرد کاربران، در آزمون بررسی میانگین، ابتدا پارامتری یا ناپارامتری بودن معیارها مورد بررسی قرار گرفته است. معیارهای زمان انجام و زمان اصلاح خطا از نوع متغیر نسبی هستند. بنابراین، ابتدا آزمون‌های بررسی نرمال بودن (آزمون کولموگروف-اسمیرنوف) بر روی آن‌ها اعمال شده است. معیارهای تعداد گام و تعداد خطا از نوع متغیر گسسته و انجام موفق از نوع متغیر ترتیبی است. بنابراین، برای تحلیل آن‌ها از آزمون ناپارامتری «مان-ویتنی» استفاده شده است.

تحلیل دیگری که در این پژوهش انجام شده، بررسی همبستگی عملکرد کاربران در تعامل با نرم‌افزار با پاسخ‌های آن‌ها در پرسشنامه بوده است. پاسخ‌های پرسشنامه از نوع رتبه‌ای است و متغیرهای عملکرد از نوع متغیرهای پیوسته، گسسته و رتبه‌ای هستند. در نتیجه، برای تحلیل این نتایج از آزمون همبستگی «اسپیرمن» استفاده شده است.

از آنجا که پژوهش حاضر بر اساس آزمایش تجربی اجرا شده، در آن مخاطرات مختلفی که یک ارزیابی تجربی را از جنبه ساختار، ورودی، خروجی و نتایج تحت تأثیر قرار می‌دهد (Wohlin et al. 2012; Grigera et al. 2016) مورد توجه قرار گرفته است.

به‌عنوان نمونه، مخاطره کم بودن اندازه نمونه با استفاده از محاسبه تعداد نمونه حداقلی و با استفاده از روش «کوهن» لحاظ شده است (Faul et al. 2007). قابلیت اطمینان رفتار و اتفاقات تصادفی در حین آزمایش با اجرای آزمون تجربی در یک محیط یکسان با سیستم‌های کامپیوتری مشابه، با شرایط تحت کنترل در نظر گرفته شده است. برای جلوگیری از مخاطرات خستگی و یادگیری، آزمایش در بازه زمانی کمتر از ۲۰ دقیقه برای هر فرد انجام شده و هر کاربر به صورت تصادفی فقط از یک نسخه نرم‌افزار استفاده

کرده است. معیارهای ارزیابی به صورت کمی و قابل اندازه‌گیری به صورت خود کار و پرسشنامه ارزیابی بر اساس تحقیقات مشابه انتخاب شده است (Davis and Davis 1989; Hornbæk 2006; Gorgian et al 2016).

۴. نتایج ارزیابی قابلیت استفاده در کارکردهای مختلف

در این بخش، نتایج ارزیابی برای هر یک از کارکردهای مورد نظر در آزمایش به تفکیک ارائه می‌گردد. همچنین، نتایج ارزیابی با استفاده از پرسشنامه با نتایج عملکرد کاربر بر اساس آزمون‌های آماری مقایسه می‌گردد.

۴-۱. کارکرد جست‌وجوی کالا

در نسخه ۱ نرم‌افزار، جست‌وجوی کالا به صورت وارد کردن نام کالا و انتخاب گزینه جست‌وجو انجام می‌شود. در نسخه ۲ نرم‌افزار، همان کارکرد جست‌وجوی کالا به اضافه قابلیت تکمیل خودکار طراحی و پیاده‌سازی شده است. نتایج ارزیابی این قابلیت با استفاده از پرسشنامه و معیارهای ارزیابی تجربه کاربری در جدول ۲ ارائه شده است. چنانچه نتایج این جدول نشان می‌دهد، در نسخه ۱ نرم‌افزار از ۲۴ شرکت کننده، ۱۶ نفر از قابلیت جست‌وجو استفاده کرده‌اند و سایر افراد، از روشی غیر از جست‌وجو برای پیدا کردن کالای مورد نظر خود استفاده کرده‌اند. در نسخه ۲ نرم‌افزار، که جست‌وجو با قابلیت تکمیل خودکار بوده است، از تعداد ۲۴ شرکت کننده، فقط ۱۰ نفر از کارکرد جست‌وجوی کالا به همراه تکمیل خودکار استفاده کرده‌اند.

مقایسه نتایج ارزیابی در دو نسخه نرم‌افزار برای این کارکرد نشان می‌دهد که مقادیر ارزیابی بر اساس عملکرد در نسخه ۱ و ۲ نرم‌افزار تفاوت معناداری نداشته است. نتایج ارزیابی از طریق پرسشنامه میان دو نسخه از نرم‌افزار، اختلاف معناداری را نشان می‌دهد. به عبارت دیگر، در حالی که زمان انجام جست‌وجوی کالا در نسخه جدید کاهش نیافته، تعداد تلاش‌های کاربران نیز از نظر تعداد گام تفاوتی نداشته است، اما به صورت میانگین از نظر آن‌ها سهولت استفاده بهبود یافته است.

جدول ۳. بررسی مقایسه‌ای نتایج ارزیابی در کارکرد جست‌وجوی کالا

ارزیابی بر اساس پرسشنامه	معیارهای ارزیابی بر اساس عملکرد					
	تعداد کل کاربران	تعداد کاربران استفاده‌کننده از کارکرد	انجام موفق	زمان انجام	تعداد کام	نتایج پرسشنامه
نسخه ۱	۲۴	۱۶	۱۶	۱۷/۳۶۵	۱/۱۵۶	۴/۱۸۸
نسخه ۲	۲۴	۱۰	۱۰	۱۷/۳۲۵	۱/۱۶۷	۴/۹
ضریب تأثیر	۰	-	۰/۱۹۲	۰/۲۱۵	۰/۸۰۴	۰/۰۳۶
معناداری سطح اختلاف	-	-	ندارد	ندارد	ندارد	دارد

۴-۲. کارکرد بازگرداندن کالا به سبد خرید

تفاوت دو نسخه نرم‌افزار در این کارکرد امکان خنثی کردن عملیات است. بدین منظور در نسخه ۱، این قابلیت به صورت مستقیم در نرم‌افزار تعبیه نشده است. بنابراین، مثلاً برای خنثی کردن عملیات حذف کالا از سبد خرید لازم است کاربر کالای حذف شده را به سبد اضافه نماید. در این نسخه، اضافه کردن کالا نیازمند جست‌وجو و سپس، اضافه کردن آن کالا است، در حالی که در نسخه ۲، با انتخاب دکمه خنثی‌سازی این کارکرد در اختیار کاربر قرار داده شده است.

نتایج ارزیابی این عملیات در دو نسخه نرم‌افزار در جدول ۳، ارائه شده است. مقایسه نتایج ارزیابی در دو نسخه نرم‌افزار برای این کارکرد نشان می‌دهد که در ارزیابی بر اساس عملکرد و ارزیابی بر اساس پرسشنامه در میان مقادیر تفاوت قابل توجهی وجود ندارد؛ بدین معنا که از نظر معیارهای عملکرد کاربران در دو نسخه اختلاف معناداری وجود دارد. تحلیل آماری نتایج پرسشنامه نیز میان دو نسخه از نرم‌افزار اختلاف معناداری را نشان می‌دهد.

جدول ۴. بررسی مقایسه‌ای نتایج ارزیابی در کارکرد بازگرداندن کالا به سبد خرید

ارزیابی بر اساس پرسشنامه	معیارهای ارزیابی بر اساس عملکرد					
	تعداد گام	زمان انجام	انجام موفق	تعداد کاربران استفاده‌کننده از کارکرد	تعداد کل کاربران	نسخه
نتایج پرسشنامه	۴/۶۱۹	۳۱/۸۱	۲۱	۲۴	۲۴	نسخه ۱
	۲/۱۲۵	۱۱	۱۶	۲۴	۲۴	نسخه ۲
	۰	۰	۰/۲۳۲	-	-	ضریب تأثیر
	دارد	دارد	ندارد	-	-	معناداری سطح اختلاف

۳-۴. کارکرد خرید کالای مشاهده‌شده

در این کارکرد از کاربران خواسته شده کالاهایی را که قبلاً بررسی و مشاهده کرده‌اند، به سبد کالای خود اضافه کنند. تفاوت دو نسخه نرم‌افزار در این کارکرد، ثبت و نمایش تاریخچه کالاهای مشاهده‌شده توسط کاربر است. در نسخه ۱، این قابلیت به صورت مستقیم در نرم‌افزار تعبیه نشده، در حالی که در نسخه ۲ کاربر شش کالایی را که به تازگی بررسی کرده در لیست تاریخچه کالاهای خود می‌بیند.

نتایج ارزیابی این عملیات در دو نسخه نرم‌افزار در جدول ۴، ارائه شده است. مقایسه نتایج ارزیابی در دو نسخه نرم‌افزار برای این کارکرد نشان می‌دهد که در ارزیابی بر اساس عملکرد زمان انجام کار و تعداد گام در انجام کار در دو نسخه نرم‌افزار اختلاف معناداری وجود دارد. نتایج نظرسنجی نیز به همین صورت است. یعنی بین نظرات افراد شرکت‌کننده در نسخه ۱ و ۲ در پاسخ سؤال به نظرسنجی اختلاف معناداری وجود دارد. به عبارت دیگر، قراردادن تاریخچه کالاهای مشاهده‌شده به صورتی معنادار موجب سهولت افزودن آن به سبد کالا شده است. این نتیجه، هم بر اساس اظهارات کاربران در نظرسنجی و هم بر اساس عملکرد آن‌ها قابل مشاهده است.

جدول ۵. بررسی مقایسه‌ای نتایج ارزیابی در کارکرد خرید کالای مشاهده‌شده

ارزیابی بر اساس پرسشنامه	معیارهای ارزیابی بر اساس عملکرد					تعداد کل کاربران	تعداد کاربرانی استفاده‌کننده از کارکرد
	نتایج پرسشنامه	تعداد گام	زمان انجام	انجام موفق	تعداد کاربران		
۲/۴۷۶	۴/۳۸۱	۴۹/۷۶۲	۲۳	۲۴	۲۴	نسخه ۱	
۴/۶۳۲	۳/۰۵۳	۱۷/۶۳۲	۱۹	۲۴	۲۴	نسخه ۲	
	۰	۰	۰/۱۹۳	-	-	ضریب تأثیر	
دارد	دارد	دارد	ندارد	-	-	معناداری سطح اختلاف	

۴-۴. کارکرد حذف کالا از سبد خرید

در این بخش از آزمایش از کاربران خواسته شده که تعدادی از کالاهای خود را از سبد خرید حذف نمایند. در نسخه اول نرم‌افزار، کاربر فقط می‌تواند هر یک از کالاها را به صورت تک به تک حذف نماید. در نسخه دوم، این امکان نیز در اختیار کاربر قرار داده شده است که چند کالا را انتخاب نموده و به صورت گروهی از سبد کالای خود حذف کند.

نتایج ارزیابی این عملیات در دو نسخه نرم‌افزار در جدول ۵، ارائه شده است. مقایسه نتایج ارزیابی در دو نسخه نرم‌افزار برای این کارکرد نشان می‌دهد که در ارزیابی بر اساس عملکرد و ارزیابی بر اساس پرسشنامه میان مقادیر تفاوت قابل توجهی وجود دارد. در حالی که نتایج نظرسنجی بسیار نزدیک است (۴/۴ در برابر ۴/۶۵)، نتایج ارزیابی آماری هم نشان می‌دهد که نتایج دو نظرسنجی اختلاف معناداری ندارند. اما در عملکرد کاربران می‌توان اختلاف قابل توجهی را مشاهده نمود (نتایج آزمون «مان-ویتنی» برای هر دو معیار زمان انجام و تعداد گام، اختلاف معناداری را نشان می‌دهد).

جدول ۶. بررسی مقایسه‌ای نتایج ارزیابی در کارکرد حذف کالاها از سبد خرید

ارزیابی بر اساس پرسشنامه	معیارهای ارزیابی بر اساس عملکرد					نتایج پرسشنامه
	تعداد کاربران	تعداد کاربران استفاده‌کننده از کارکرد	انجام موفق	زمان انجام	تعداد کام	
نسخه ۱	۲۴	۲۴	۲۴	۶/۵۲۲	۱۳/۶۹۶	۴/۵۶۵
نسخه ۲	۲۴	۲۲	۲۲	۳/۸۴۵	۸/۸	۴/۴
ضریب تأثیر	-	-	۰,۴۸۸	۰,۰	۰,۰	۰/۹۷۷
معناداری سطح اختلاف	-	-	ندارد	دارد	دارد	ندارد

۴-۵. کارکرد تکمیل اطلاعات

در این کارکرد از کاربران خواسته شده که تعدادی فیلد اطلاعاتی (مرتبط با نام، و نام خانوادگی و ...) را تکمیل نمایند. در نسخه ۱ نرم‌افزار، در زمان ورود اطلاعات اعتبارسنجی صورت نمی‌گیرد و تنها در زمان ثبت خطاهای داده ورودی به کاربر اعلام می‌شود؛ در حالی که در نسخه ۲ نرم‌افزار، در زمان تکمیل اطلاعات اعتبارسنجی انجام می‌شود و به کاربر پیغام خطای مناسب به‌منظور رفع خطا اعلام می‌گردد.

این کارکرد ابتدا با ۲۴ شرکت‌کننده انجام شده است. اما بررسی انحراف معیار اشتراکی و ضریب تأثیر نتایج ارزیابی نشان می‌دهد که تفاوت میانگین دو نسخه نرم‌افزار ممکن است به‌علت کوچک بودن اندازه نمونه باشد (در حالی که در بقیه کارکردها این مشکل وجود ندارد). به‌همین علت، با استفاده از ابزار "G*Power" تعداد نمونه کافی برای ارزیابی محاسبه شده است. این مقدار در کارکرد تکمیل اطلاعات معادل ۳۵ نفر است. به‌همین علت، مجدداً آزمایش به‌صورت کامل تکرار شده تا تعداد نمونه کافی برای اعتبار ارزیابی وجود داشته باشد. به‌همین علت، تعداد شرکت‌کنندگان در این کارکرد ۳۵ است که با سایر کارکردها متفاوت است. از بین این ۳۵ نفر در هر دو نسخه نرم‌افزار، ۳۵ نفر موفق به تکمیل اطلاعات شده‌اند. نتایج ارزیابی در جدول ۷، نشان داده شده است. بررسی‌های آماری نشان می‌دهد که دو نسخه از نظر معیارهای عملکرد کاربران اختلاف معناداری دارد، در حالی که مطابق با نظرسنجی سهولت استفاده از نرم‌افزار در دو نسخه تفاوت معناداری وجود نداشته است.

جدول ۷. بررسی مقایسه‌ای نتایج ارزیابی در کارکرد تکمیل اطلاعات

ارزیابی بر اساس پرسشنامه	معیارهای ارزیابی بر اساس عملکرد							
	تعداد کل کاربران	تعداد استفاده‌کننده از کارکرد	تعداد کاربران موفق	زمان انجام	تعداد کام	تعداد خطا	زمان اصلاح خطا	نتایج پرسشنامه
نسخه ۱	۳۵	۳۵	۳۵	۱۲۵/۷۱۴	۳/۶	۴/۰۹	۷۱/۲	۴
نسخه ۲	۳۵	۳۵	۳۵	۹۸/۳۱۴	۱/۳۴	۲/۵۷	۴۸/۷۲۷	۴/۲۶۷
ضریب تأثیر	-	-	۱/۰۰	۰/۰۰۸	۰۰۰	۰	۰/۰۰۶	۰/۰۱۸۲
معناداری سطح اختلاف	-	-	ندارد	دارد	دارد	دارد	دارد	ندارد

۵. همبستگی بین پرسشنامه و تعاملات کاربران

در این قسمت از مقاله، در راستای مقایسه نتایج ارزیابی از طریق معیارهای مبتنی بر عملکرد کاربر و پرسشنامه، به بررسی عملکرد هر یک از کاربران می‌پردازیم. در این راستا، به مجموعه سؤالات زیر که در این پژوهش مطرح شده، پاسخ خواهیم داد:

- ◇ آیا هنگامی که کاربر زمان بیشتری را صرف انجام کاری در نرم‌افزار می‌کند، در پرسشنامه به سهولت استفاده آن کارکرد امتیاز کمتری می‌دهد؟
- ◇ آیا هنگامی که کاربر گام‌های بیشتری طی می‌کند تا کاری را در نرم‌افزار انجام دهد، در پرسشنامه به سهولت استفاده آن کارکرد امتیاز کمتری می‌دهد؟
- ◇ آیا هنگامی که کاربر در انجام یک کار به خطاهای بیشتری برخورد می‌کند، در پرسشنامه به سهولت استفاده آن کارکرد امتیاز کمتری می‌دهد؟
- ◇ آیا هنگامی که کاربر در انجام یک کار زمان بیشتری را صرف برطرف کردن خطاها می‌کند، در پرسشنامه به سهولت استفاده آن کارکرد امتیاز کمتری می‌دهد؟

به‌منظور پاسخ به سؤالات فوق، از آزمون همبستگی معیارها استفاده می‌کنیم. با استفاده از این آزمون می‌توان دریافت که آیا بین تعاملات ثبت‌شده هر کاربر با جوابی که در پرسشنامه وارد می‌کند، رابطه و همبستگی وجود دارد؟ بدین منظور، انتخاب آزمون آماری مناسب حائز اهمیت است. از آنجا که معیار سهولت استفاده از نوع متغیر ترتیبی است و باقی معیارها از نوع متغیرهای نسبی، گسسته و ترتیبی هستند، برای محاسبه

همبستگی میان معیار سهولت استفاده و سایر معیارها، ضریب همبستگی رتبه‌ای «اسپیرمن» مورد استفاده قرار می‌گیرد.

۵-۱. همبستگی پرسشنامه و عملکرد کاربران موفق

در این بررسی، ارزیابی بر روی هر کارکرد به صورت جداگانه انجام شده است. به عنوان نمونه، برای کارکرد جست‌وجوی کالا، همه کاربرانی که در نسخه ۱ و ۲ از قابلیت جست‌وجوی کالا در نرم‌افزار استفاده کرده‌اند (در مجموع، ۴۸ شرکت کننده) و جست‌وجوی خواسته شده را با موفقیت انجام داده‌اند، مورد بررسی قرار گرفته‌اند. سپس، آزمون همبستگی رتبه‌ای «اسپیرمن» بر روی داده‌های معیارهای عملکرد (مانند زمان انجام کار و تعداد گام‌ها در انجام هر کار) و پرسشنامه اعمال شده است. نتایج این ارزیابی در جدول ۷، نشان داده شده است.

بنا بر نتایج ارزیابی، تنها در دو کارکرد بازگرداندن کالا به سبد خرید و خرید کالای مشاهده شده (که دارای P-value کمتر از ۰/۰۵ هستند)، همبستگی معناداری مشاهده شده است. در کارکرد بازگرداندن کالا به سبد خرید شدت رابطه بین معیار سهولت استفاده و زمان انجام کار برابر ۰/۶۲۳- و بین معیار سهولت استفاده و تعداد گام‌ها برابر ۰/۵۷۵- است. یعنی در این کارکرد، کاربرانی که زمان انجام کار کمتری داشته‌اند و تعداد گام‌های کمتری برای انجام آن طی کرده‌اند، به سهولت استفاده کارکرد بازگرداندن کالا به سبد خرید امتیاز بهتری در پرسشنامه داده‌اند. این رابطه در کارکرد خرید کالای مشاهده شده نیز نمایان است. همبستگی بین معیار سهولت استفاده و دو معیار زمان انجام و تعداد گام در این کارکرد، به ترتیب دارای اندازه متوسط ۰/۵۸۸- و ۰/۵۵۰- است.

همان‌طور که در جدول ۹، مشاهده می‌شود، در سه کارکرد جست‌وجوی کالا، بازگرداندن کالا به سبد خرید و تکمیل اطلاعات، سطح معناداری آزمون «اسپیرمن» بین معیار سهولت استفاده و دو معیار کمی بیشتر از ۰/۰۵ است. این مقادیر نشان می‌دهند که در کارکردهای مذکور رابطه همبستگی بین داده‌های زمان انجام کار و پرسشنامه یا تعداد گام‌ها و پرسشنامه وجود ندارد. در نتیجه، بین عملکرد کاربر و نتایج پرسشنامه همبستگی وجود ندارد.

همچنین، سطح معناداری بین معیار سهولت استفاده و دو معیار تعداد خطا و زمان اصلاح خطاها در کارکرد تکمیل اطلاعات به ترتیب برابر ۰/۲۶۳ و ۰/۳۲۶ بوده و به این

معناست که بین پرسشنامه و این معیارها نیز همبستگی مشاهده نشده است.

جدول ۸. بررسی همبستگی معیارهای ارزیابی مبتنی بر عملکرد و پرسشنامه

کارکرد	کاربران تعداد کل	موفقیت کاربران تعداد	همبستگی زمان انجام کار و پرسشنامه		همبستگی تعداد گام‌ها و پرسشنامه		همبستگی زمان اصلاح خطا و پرسشنامه	
			P-value	همبستگی رگرسیونی	P-value	همبستگی رگرسیونی	P-value	همبستگی رگرسیونی
جست‌وجوی کالا	۴۸	۳۲	۰/۰۶۲۸	۰/۰۸۹	۰/۳۱۱	۰/۱۸۵		
بازگرداندن کالا به سید خرید	۴۸	۴۳	۰	-۰/۶۲۳	۰	-۰/۵۷۵		
خرید کالای مشاهده‌شده	۴۸	۴۴	۰	-۰/۵۸۸	۰	-۰/۵۵۰		
حذف کالاها از سید خرید	۴۸	۴۳	۰/۹۳۴	۰/۰۱۳	۰/۹۳۱	-۰/۰۱۴		
تکمیل اطلاعات	۷۰	۷۰	۰/۷۱۷	-۰/۰۴۴	۰/۱۱۴	-۰/۱۹۱	۰/۳۲۶	-۰/۱۲۱
مجموع	۱۹۲	۱۶۲						

۲-۵. همبستگی پرسشنامه و عملکرد کاربران ناموفق

در این پژوهش، پنج کارکرد مختلف توسط کاربران مختلف در نرم‌افزار مورد استفاده قرار گرفته است. در مجموع، ۱۹۲ (نفر/کارکرد) مورد توسط کاربران مختلف اجرا و ارزیابی شده است. از این بین، ۱۶۲ مورد به‌صورتی موفقیت‌آمیز اجرا و ۳۰ مورد اجرا نشده است. جهت دقت نتایج ارزیابی، کاربران ناموفق در نتایج ارزیابی در نظر گرفته نشده‌اند. اما توجه به عملکرد کاربران ناموفق، می‌تواند روش پاسخ‌دهی به پرسشنامه را از نظر دقت پاسخ‌ها برای ما مشخص نماید. به‌همین علت، بررسی دیگری که در این پژوهش انجام شده، همبستگی داده‌های پرسشنامه و عملکرد کاربرانی است که کارکردهای مورد نظر را انجام نداده‌اند.

در مواردی که کارکردی اجرا نشده است، ممکن است کاربران نتوانسته باشند که کارکرد را به انجام برسانند (انجام کار برای آن‌ها به حدی سخت بوده که نتوانند آن را انجام دهند) یا تمایلی به انجام آن نداشته‌اند.

در آزمایش‌های انجام‌شده در این پژوهش، ۸۱/۲۵ درصد افراد ناموفق به‌سهولت

استفاده در پرسشنامه خود عدد بالاتر از ۴ داده‌اند. این نتیجه نشان می‌دهد که این کاربران در تکمیل اطلاعات پرسشنامه (در کارکردهای ناموفق) بدون توجه به عملکرد خود، سؤالات مربوط به پرسشنامه را تکمیل کرده‌اند.

۶. بحث و تفسیر نتایج

در این بخش از مقاله، تحلیل و تفسیر نتایج حاصل از این پژوهش ارائه می‌گردد. لازم به ذکر است که این نتایج در محدوده پژوهش انجام‌شده قابل تفسیر و استناد است. آزمون‌های آماری کفایت داده‌ها برای نتایج حاصل را در نتایج گزارش شده تأیید می‌کنند. اما محدودیت‌هایی در این پژوهش (مشابه هر آزمون تجربی دیگری) وجود دارد. آزمایش طراحی شده در دو نسخه از یک نرم‌افزار، با پنج کارکرد و پنج معیار اجرا شده است.

۶-۱. عدم همبستگی نتایج پرسشنامه و معیارهای عملکرد کاربران

آزمایش‌های انجام‌شده در این پژوهش نشان می‌دهد که میان نتایج حاصل از ارزیابی با استفاده از پرسشنامه کاربری (سهولت استفاده) و نتایج ارزیابی بر اساس تجربیات کاربری (معیارهای انجام موفق، زمان انجام، تعداد گام، تعداد خطا، زمان اصلاح خطا) در نسخه‌های مختلف یک نرم‌افزار همبستگی وجود ندارد. به عبارت دیگر، اگر کاربران نسخه‌ای جدید از نرم‌افزار، سهولت استفاده آن را بیشتر (یا کمتر) از نسخه قبل ارزیابی کنند، لزوماً به این معنا نیست که زمان انجام کار توسط کاربران کاهش (افزایش) می‌یابد، تعداد گام‌های کمتر (بیشتر) برای انجام کار برمی‌دارند، خطاهای کمتری در انجام کار با نرم‌افزار خواهند داشت، یا در زمان کمتری خطاهای خود را برطرف می‌کنند.

از طرف دیگر، اگر تغییرات در نرم‌افزار منجر به نسخه‌ای جدید شود که کاربران بتوانند یک کارکرد را در زمان کمتری انجام دهند، تعداد گام‌ها خطای کمتری در انجام آن داشته باشند، خطاهای کمتری انجام دهند یا سریع‌تر بتوانند خطاها را برطرف کنند، لزوماً به این معنا نیست که در ارزیابی نسخه جدید از نرم‌افزار به سهولت استفاده امتیاز بیشتری خواهند داد.

این نتایج اگرچه در ابتدا ناامیدکننده به نظر می‌رسند، اما در انتخاب روش ارزیابی قابلیت استفاده، اطلاعات قابل توجهی ارائه می‌دهند. بدین معنا که برای ارزیابی سهولت

استفاده یک نرم‌افزار لازم است ابتدا به صورت دقیق معیاری که مرتبط با سهولت استفاده است، تعریف گردد و سپس، در فرایند ارزیابی اندازه‌گیری شود. پرسشنامه لزوماً اطلاعات دقیقی از نظرات کاربران (و معیارهایی که برای بهبود باید بر آن تکیه کنیم)، برای ما ایجاد نمی‌کند. اما اگر می‌خواهیم به صورت کلی نظرات کاربران را از نظر سهولت استفاده یک کارکرد بدانیم، روش پرسشنامه می‌تواند روش مناسب‌تری باشد؛ زیرا ممکن است لزوماً معیارهایی که ما به عنوان سهولت استفاده در نظر می‌گیریم، معیارهای مورد نظر کاربر برای سهولت استفاده نباشد و سایر عوامل که برای ما ناشناخته است، بر ارزیابی کاربر اثرگذار باشد.

یکی از موضوعاتی که به عنوان پژوهش آینده می‌تواند تعریف شود، تعیین معیارهای سهولت استفاده از نظر کاربران است. در این پژوهش، ما معیارهای تعریف شده منطبق با سهولت استفاده در پژوهش‌های مرتبط (Hornbæk (2006) و Davis and Davis (1989) را مبنا قرار داده‌ایم. انتخاب سایر معیارها و انجام آزمایش مشابه می‌تواند در یافتن معیارهای سهولت استفاده از دید کاربران یا تعمیم عدم همبستگی به صورت عمومی راهگشا باشد.

۶-۲. کاربران ناموفق حفظ ظاهر می‌کنند

بررسی‌های ما نشان می‌دهد که بیش از ۸۰ درصد از کاربرانی که در کارکردهای مختلف موفق به اجرای کارکرد نشده‌اند، به سهولت استفاده کارکرد، امتیازی بین «خوب» یا «بسیار خوب» داده‌اند. این نتایج نظرات «استفن - دیویدویتز»^۱ را در ارتباط با افرادی که در نظرسنجی شرکت می‌کنند، تأیید می‌کند. ممکن است افراد به دلایل مختلف، نظراتی اعلام کنند که با نظر واقعی یا عملکرد واقعی آن‌ها در ارتباط با نرم‌افزار متفاوت باشد؛ مثلاً به این علت که می‌خواهند بهتر از آنچه که هستند به نظر برسند.

بنا بر نتایج این بررسی، عدم اظهار نظرات واقعی یکی از چالش‌های ارزیابی قابلیت استفاده با استفاده از پرسشنامه است. در نتیجه، در ارزیابی‌هایی که فقط بر اساس پرسشنامه انجام می‌شود، این مسئله می‌تواند منجر به خدشه‌دار شدن نتایج ارزیابی شود. چون در ارزیابی‌هایی که فقط مبتنی بر پرسشنامه است، ما ابزاری برای کنترل موفقیت یا عدم موفقیت کاربران - و در نتیجه حذف نظرات کاربران ناموفق - نخواهیم داشت. در این

1. Stephens-Davidowitz

پژوهش از آنجا که ما از هر دو روش به صورت همزمان استفاده کرده‌ایم، این امکان را داریم که کاربران ناموفق را شناسایی کرده و پرسشنامه آن‌ها را از نتایج ارزیابی خارج کنیم. بدیهی است که در نظر گرفتن این افراد، ممکن است منجر به نتایج متفاوتی شود که اعتبار کل پژوهش را با چالش مواجه کند.

۷. نتیجه‌گیری و کارهای آینده

روش‌های مختلفی برای ارزیابی قابلیت استفاده نرم‌افزار وجود دارد. این روش‌ها از نظر زمان، هزینه و دقت ارزیابی با یکدیگر متفاوت‌اند. یکی از مسائل مهم در ارزیابی قابلیت استفاده، انتخاب روش درست ارزیابی است که بتواند اطلاعاتی مفید متناسب با زمان و هزینه صرف شده برای آن در اختیار تولیدکننده نرم‌افزار قرار دهد. در این پژوهش، دو روش مختلف ارزیابی، یکی روش ارزیابی بر اساس پرسشنامه، به عنوان یک روش متداول، و ارزیابی بر اساس عملکرد کاربران در تعامل با نرم‌افزار مورد استفاده قرار گرفته و مقایسه شده است.

نتایج حاصل از مقایسه دو روش ارزیابی نشان می‌دهد که آنچه که کاربران در نظرسنجی مبتنی بر پرسشنامه در ارتباط با سهولت استفاده نرم‌افزار اعلام می‌کنند، به صورت کلی با نتایج حاصل از عملکرد آن‌ها در زمان استفاده از نرم‌افزار، از نظر انجام موفق، زمان انجام یک کارکرد، تعداد گام‌ها در اجرای یک کارکرد، تعداد خطا و زمان صرف شده برای برطرف کردن خطاها همبستگی ندارد. بدین ترتیب، یا باید معیارهای دیگری برای سهولت استفاده تعریف کرد یا نتایج حاصل از پرسشنامه را مورد تردید قرار داد. به صورت دقیق‌تر، ممکن است در دو نسخه از یک نرم‌افزار، عملکرد کاربران از نظر معیارهای مذکور به صورت معناداری بهبود یافته باشد، اما در نظرسنجی از کاربران، نظرات آن‌ها نسبت به نسخه قبلی بهبود معناداری نشان ندهد (مانند کارکرد حذف کالاها از سبد خرید و تکمیل اطلاعات). همچنین، ممکن است در نظرسنجی کاربران سهولت استفاده از یک کارکرد در دو نسخه مختلف، اختلاف معناداری مشخص شود، اما در معیارهای عملکرد کاربران، تغییر معناداری مشاهده نشود (مانند کارکرد جست‌وجوی کالا). نتایج این آزمایش همچنین نشان می‌دهد که در برخی موارد نیز این نتایج با یکدیگر همبستگی دارند (مانند کارکردهای بازگرداندن کالا به سبد خرید و خرید کالای مشاهده‌شده).

طبق بررسی‌های این پژوهش، کاربرانی که به هر دلیلی یک کارکرد را اجرا

نکرده‌اند و در اجرای آن موفق نبوده‌اند، در بالاتر از ۸۰ درصد موارد، سهولت استفاده را در حد خوب و بسیار خوب ارزیابی کرده‌اند. در این موارد، صداقت کاربران در تکمیل اطلاعات پرسشنامه مورد تردید خواهد بود که نتایج استفاده از این روش را تهدید می‌کند. بدیهی است که نتایج این پژوهش، مشابه هر آزمایش تجربی دیگری باید در محدوده آزمایش تفسیر گردد. هرچند، آزمون‌های آماری کفایت داده‌ها برای نتایج حاصل را در این آزمایش تأیید می‌کنند، اما این مقاله قصد تعمیم آن به همه کاربران و نرم‌افزارهای کاربردی را ندارد و برای تعمیم آن نیاز به اجرای آزمایش‌های بیشتر در نرم‌افزارهای مختلف با معیارهای متنوع‌تر برای ارزیابی عملکرد کاربران متنوع وجود دارد. این آزمایش‌ها می‌تواند به‌عنوان پژوهش‌های آینده در این مسیر در نظر گرفته شود. در این راستا می‌توان از سایر معیارهای ارزیابی کمی قابلیت استفاده در طراحی آزمایش استفاده کرد. در این شرایط، می‌توان همبستگی میان سهولت استفاده با سایر معیارها را سنجید و در نهایت، به این پرسش پاسخ داد که آیا سهولت استفاده از دیدگاه کاربران با معیار یا معیارهای مشخصی از قابلیت استفاده همبستگی دارد یا خیر.

بررسی‌های ما در این پژوهش نشان می‌دهد که نتایج حاصل از پرسشنامه در کارکردهای بازگرداندن کالا به سبد خرید و خرید کالای مشاهده‌شده، با نتایج حاصل از عملکرد همخوانی دارد. اما در کارکردهای حذف کالا، جست‌وجوی کالا و تکمیل اطلاعات این نتایج همخوانی ندارد. بررسی سطح پیچیدگی کارکرد با این نتایج نشان می‌دهد که در کارکردهایی که پیچیدگی بیشتری دارند، این همبستگی وجود دارد. اما برای کارکردهایی که ساده هستند، یا کاربران به‌صورت متداول با آن‌ها در نرم‌افزارهای مختلف کار می‌کنند، این همبستگی وجود ندارد. بنابراین، ممکن است سطح پیچیدگی کارکرد یکی از معیارهایی باشد که در نحوه پاسخ‌دهی به پرسشنامه مؤثر باشد. ارزیابی دقیق این موضوع نیاز به پژوهش‌های بیشتری دارد که می‌تواند به‌عنوان پژوهش‌های آینده در نظر گرفته شود. همچنین، می‌توان بررسی کرد که آیا سهولت استفاده از دیدگاه کاربران به معیارهای دیگری بجز معیار عملکرد آن‌ها (مانند پیچیدگی کارکرد، نوع و تنوع قابلیت‌هایی که در اختیار آن‌ها قرار می‌گیرد و ...) همبستگی دارد؟

فهرست منابع

آخسیک، سمیه‌سادات، و رضا رجبعلی‌بگلو. ۱۳۹۲. استفاده‌پذیری نرم‌افزار سیمرغ از دید کاربران: لزوم

- بازنگری در پارادایم طراحی نرم افزارهای کتابخانه‌ای. فصلنامه کتابداری و اطلاع‌رسانی ۱۶ (۲): ۲۹-۴۷.
- آقارضایی، ژیلا، رضا خواجهویی، لیلا احمدیان، و لاله آقارضایی. ۱۳۹۲. ارزیابی کاربردپذیری سیستم اطلاعات آزمایشگاه. نشریه مدیریت اطلاعات سلامت ۱۰ (۲): ۲۱۳-۲۲۴.
- ابراهیمی، ابوالقاسم، نیلوفر ایمان‌خان، و عبدالرضا اسماعیلی. ۱۳۹۴. ارائه مدلی برای تبیین عوامل مؤثر بر تمایلات رفتاری کاربران وبسایت‌های خبرگزاری اینترنتی (مورد مطالعه: دانشجویان دانشگاه سمنان). نشریه مدیریت فناوری اطلاعات ۷ (۳): ۴۷۳-۴۹۲.
- احمدیان، لیلا، فاطمه صالحی، عاطفه عابدین‌زاده، و فاطمه خطیبی. ۱۳۹۶. بررسی کاربردپذیری سیستم اطلاعات رادیولوژی. نشریه مدیریت سلامت ۶۷-۷۸.
- بابایی، حسین، محمود صالح اصفهانی، و سیده فاطمه نورانی. ۱۳۹۲. چارچوبی برای شناسایی و رتبه‌بندی شاخص‌های ارزیابی کیفیت خدمات الکترونیک (مورد کاوی: وبسایت گلستان دانشگاه پیام نور). فصلنامه علمی-پژوهشی فناوری اطلاعات و ارتباطات ایران ۵ (۱۷ و ۱۸): ۱-۲۲.
- پدرام، حسین، داود کریم‌زادگان‌مقدم، و ژاله عاشقی. ۱۳۹۰. به کارگیری استاندارد ایزو ۹۱۲۶ در ارزیابی کیفیت سامانه‌های یادگیری الکترونیکی در ایران (مطالعه موردی: سامانه آموزش مجازی دانشگاه صنعتی امیرکبیر). پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۲۷ (۲): ۴۹۵-۵۱۷.
- تهرانی، رضا، و حمید جمشیدی. ۱۳۹۴. تحلیل میزان تأثیرگذاری عوامل اطلاعاتی، سیستمی و طراحی رابط کاربر بر وفاداری مشتریان وبسایت‌ها با توجه به نقش رضایت و اعتماد. پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۳۰ (۴): ۱۰۸۵-۱۱۰۶.
- حسن‌زاده، محمد، و فائزه اسکندری. ۱۳۹۰. ارزیابی رابط کاربر نرم افزارهای علوم اسلامی شهر قم. پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۲۶ (۳): ۵۹۵-۶۲۴.
- حسینی، صفیه، عارفه شریف، و محسن نوکارتیزی. ۱۳۹۵. اعتبارسنجی و ارزیابی عناصر محیط رابط کاربر کتابخانه‌های دیجیتالی فارسی کودکان. پژوهشنامه پردازش و مدیریت اطلاعات ۳۲ (۱): ۱۹۵-۲۱۸.
- خواجهویی، رضا، امیرعباس عزیز، و علیرضا آتشی. ۱۳۹۲. بررسی کاربردپذیری سیستم اطلاعات اورژانس با استفاده از روش ارزیابی اکتشافی. نشریه مدیریت اطلاعات سلامت ۱۶ (۵۲): ۶۱-۷۲.
- زره‌ساز، محمد، و رحمت‌الله فتاحی. ۱۳۹۳. شناسایی و تحلیل امکانات و قابلیت‌های راهنمایی و کمک در نرم افزارهای کتابخانه‌های دیجیتالی ایرانی. پژوهشنامه پردازش و مدیریت اطلاعات ۲۹ (۳): ۸۱۷-۸۳۴.
- _____، و محمدرضا داورپناه. ۱۳۸۵، بررسی و تحلیل عناصر و ویژگی‌های مطرح در رابط کاربر نرم افزار سیمرغ و تعیین میزان رضایت دانشجویان دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد از تعامل با این نرم افزار. فصلنامه کتابداری و اطلاع‌رسانی ۳۶ (۴): ۱۲۷-۱۵۰.
- شکاری، محمد، و داود افشاری. ۱۳۹۳. ارزیابی کاربردپذیری سیستم‌های کروماتوگرافی گازی: کارایی، اثربخشی و رضایت‌مندی. مجله علمی پژوهشی سلامت کار ایران ۱۱ (۶): ۳۴-۴۲.
- فتاحی، رحمت‌اله، محمدحسین دیانی، و سیدمهدی حسینی. ۱۳۹۲. بررسی دیدگاه کاربران پیرامون تأثیر

ویژگی‌های تعاملی رابط کاربر پایگاه‌های اطلاعاتی بر رفتار علمی آن‌ها با تأکید بر نظریه تجربه دلپذیر. پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۲۸ (۳): ۷۱۷-۷۳۹.

فرقانی، محمدمهدی، و علی رستگار. ۱۳۹۵. مقایسه کاربردپذیری تارنماهای خبرگزاری‌های ایرنا، ایسنا و فارس از دیدگاه کاربران. مجله بهار ۱۰۲: ۲۳-۴۰.

کریم‌زادگان‌مقدم، داود، محسن خداپرست، و داود وحدت. ۱۳۹۰. ارزیابی عوامل مؤثر بر رضایت یادگیرنده الکترونیکی. فصلنامه پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۲۷ (۲): ۴۷۸-۶۴۱.

مانیان، امیر، محمد موسی‌خانی، و سمانه رحیمیان. ۱۳۹۳. ارائه مدل عوامل مؤثر بر رضایت کاربران سامانه جامع آموزش دانشگاه تهران. فصلنامه مطالعات مدیریت فناوری اطلاعات ۲ (۸): ۱۲۳-۱۳۸.

نارمنجی، سیدمهدی، ۱۳۹۶. ارزیابی کاربردپذیری سامانه مدیریت کتابخانه‌های عمومی کشور (سامان) بر اساس اصول دهگانه نیلسون. نشریه تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی ۸۹: ۱۹۹-۲۱۷.

نوروزی، یعقوب. ۱۳۹۰. تحلیلی بر کاربردمداری رابط کاربر در صفحات وب فارسی کتابخانه‌های دیجیتال ایران و ارائه الگوی پیشنهادی. پژوهشنامه پردازش و مدیریت اطلاعات (علوم و فناوری اطلاعات سابق) ۲۳ (۳): ۶۵۵-۶۷۴.

نوکاریزی، محسن، حوا خیرخواه، و محمدرضا کیانی. ۱۳۹۲. ارزشیابی کاربردپذیری وب‌سایت‌های پژوهشگاه علوم و فناوری اطلاعات ایران و مرکز منطقه‌ای اطلاع‌رسانی علوم و تکنولوژی شیراز از منظر دانشجویان تحصیلات تکمیلی دانشگاه شیراز. نشریه کتابداری و اطلاع‌رسانی ۱۶ (۴): ۱۰۳-۱۲۲.

هوشمند مروستی، مریم، فرهاد بخشی، و محمد ربیعی. ۱۳۹۳. ارزیابی و تست قابلیت استفاده طراحی واسط کاربر برای سیستم‌های اطلاعاتی اورژانس بیمارستان. فصلنامه بیمارستان ویژه‌نامه تله مدیسن ۴ (۱): ۲۴-۳۹.

References

- Bargas-Avila, J. A., S. Orsin, H. Piosczyk, D. Urwyler, and K. Opwis. 2011. Enhancing online forms: Use format specifications for fields with format restrictions to help respondents. *Interacting with Computers* 23 (1): 33-39.
- Carter, P. 2007. Liberating usability testing. *Interactions* 14 (2): 18-22.
- Celis-Morales, C. A., F. Perez-Bravo, L. Ibañez, C. Salas, M. E. S. Bailey, and J. M. R. Gill. 2012. Objective vs. self-reported physical activity and sedentary time: Effects of measurement method on relationships with risk biomarkers. *PLoS ONE* 7 (5): e36345.
- Davis, F. D. and F. Davis. 1989. Perceived Usefulness· Perceived Ease of Use and User Acceptance of Information Technology. *Management Information Systems Research Center University of Minnesota Stable* 13 (3): 319-340.
- Distante, Damiano, Alejandra Garrido, Julia Camelier-Carvajal, Roxana Giandini, and Gustavo Rossi. 2014. Business processes refactoring to improve usability in e-commerce applications. *Electronic Commerce Research* 14 (4): 497-529.
- Faul, Franz, Edgar Erdfelder, Albert-Georg Lan, and Axel Buchner. 2007. G* power 3: a flexible statistical

- power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39 (2): 175–191.
- Frøkjær, E., M. Hertzum, and K. Hornbæk. 2000. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. The Hague, The Netherlands: 345-352.
- Grigera, Julián, Alejandra Garrido, Damiano Distanto, and Gustavo Rossi. 2016. Assessing refactorings for usability in e-commerce applications. *Empirical Software Engineering* 21 (3): 1224-1271.
- Hal, J., J. Overgaard, and C. Grau. 2003. Evaluation of objective measures of smoking status—A prospective clinical study in a group of head and neck cancer patients treated with radiotherapy. *Acta Oncologica* 42 (2): 154–159.
- Hornbæk, K. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies* 64 (2): 79–102.
- Hornbæk, K., and E. L. Law. 2007. Meta-analysis of correlations among usability measures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California: 617-626.
- ISO (2011) *ISO/IEC 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*. Geneva: International Organisation for Standardisation, Geneva.
- Junco, R. 2013. Comparing actual and self-reported measures of facebook use. *Computers in Human Behaviour* 29: 626-631.
- Lewis, James R. 2018. Measuring Perceived Usability: The CSUQ, SUS, and UMUX. *International Journal of Human-Computer Interaction* 34 (12): 1148-1156.
- L., L., Maojuan Penga, Nan Jiangb, and Rob Lawb. 2017. An empirical study on the influence of economy hotel website quality on online booking intentions. *International Journal of Hospitality Management* 63: 1-10.
- Nathan, R. J., and P. H. P. Yeow. 2011. Crucial web usability factors of 36 industries for students: a large-scale empirical study. *Electronic Commerce Research*. 11 (2): 151-180.
- Otte, J. J., B. Littenberg, & J. R. Harvey-Berino. 2010. Relationship between selfreport and an objective measure of television-viewing time in adults. *Obesity* 18 (6): 1273–1275.
- Pauwels, S. L., C. Hübscher, S. Leuthol, J. A. Bargas-Avila, and K. Opwis. 2009. Error prevention in online forms: Use color instead of asterisks to mark required-fields. *Interacting with Computers* 21 (4): 257–262.
- Rangardt, J., and M. Czaja. 2017. Empirical investigation of how user experience is affected by response time in a web application. Master of Science thesis in software engineering. Chalmers University of technology, Sweden.
- Raza, Arif, Luiz Fernando Capretz, and Faheem Ahmed. 2012. Users' perception of open source usability: An empirical study. *Engineering with Computers* 28 (2): 109-121.
- Rodríguez, Francy D., Silvia T. Acuña, and Natalia Juristoa. 2015. Design and programming patterns for implementing usability functionalities in web applications. *Systems and Software* 105: 107-124.
- Saleh, A., R. B. Isami, and N. B. Fabil. 2015. Extension of PACMAD model for usability evaluation metrics using Goal Question Metrics (GQM) approach. *Journal of Theoretical and Applied Information Technology* 79 (1): 90-100.
- Seckle, M., S. Heinz, J. A. Bargas-Avila, K. Opwis, and A. N. Tuchk. 2014. Designing usable web forms – empirical evaluation of web form improvement guidelines. *32nd Annual ACM Conf. on Human Factors in Computing Systems*. Toronto, Canada: 1275-1284.
- Stephens-Davidowitz, Seth. 2018. *Everybody Lies: Big Data, New Data, and what the Internet can Tell*

Us about who We Really are. New York: William Morrow & Co.

Wohlin, C., P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. 2012. *Experimentation in software engineering: an introduction.* Norwell: Springer.

فاطمه مازنی

دانش آموخته کارشناسی ارشد در رشته مهندسی کامپیوتر از دانشگاه گلستان است.

مهندسی نرم افزار و ارزیابی قابلیت استفاده نرم افزارهای کاربردی از جمله علایق پژوهشی وی است.



شیوا وفادار

دارای مدرک تحصیلی دکتری در رشته مهندسی کامپیوتر از دانشگاه صنعتی امیرکبیر است. ایشان هم اکنون استادیار گروه مهندسی کامپیوتر دانشگاه گلستان است.

مهندسی نرم افزار، مهندسی کیفیت و مهندسی هوش از جمله علایق پژوهشی وی است.

