

The JUS Top 10 Articles: 2005–2021

Bill Albert

Executive Director
User Experience Center
Bentley University
Waltham, MA 02452 USA
walbert@bentley.edu

James R. (Jim) Lewis

Distinguished User
Experience Researcher
MeasuringU
jjim@measuringu.com

The [first issue of the Journal of Usability Studies](#) was published in November 2005 with [Avi Parush](#) as the Editor in Chief and an [essay by Jakob Nielsen](#).

That means that with the current issue, August 2021, JUS has completed its 16th volume of essays and peer-reviewed articles by and for usability and UX researchers and practitioners. Counting this issue, that's 71 essays by leaders in the field and 173 peer-reviewed articles.

On this occasion, we thought we'd check our server stats to find the top 10 articles and essays. As you will see, our readers' interest is diverse, including topics such as card sorting, electronic medical records, quantitative methods, and even parallax scrolling. We hope you enjoy this special essay, and you find some helpful research that may be useful to you. We also encourage you to explore our entire [inventory of articles and essays](#).



The Top Ten

There are various ways to come up with a “top-ten list.” In many ways, it is a fool’s errand since there is never any consensus. However, we wanted to acknowledge some of the outstanding articles we have published since our inception in 2005. We also wanted to share some amazing research that our readers may have missed over the years.

One obvious way to find our top articles is to look at the number of citations through Google Scholar or similar platforms. This would certainly be one way to measure the academic impact of an article. However, many of our readers simply want to read, learn, and hopefully apply these insights into their professional work. Therefore, we turned to our own analytics and simply identified the ten articles that have received the most page views. While not a perfect measure by any stretch, it does give us some indication about the level of interest.

Below we start our countdown, from 10 to 1, of the most viewed articles on the JUS website. We’ve included the abstract from each article, followed by a short commentary focused on the impact of the article.

#10: Brown, M. E., & Hocutt, D. L. (2015). [Learning to use, useful for learning: A usability study of Google Apps for Education](#). *Journal of Usability Studies*, 10(4), 160–181.

Abstract

Using results from an original survey instrument, this study examined student perceptions of how useful Google Apps for Education (GAPE) was in students’ learning of core concepts in a first-year college composition course, how difficult or easy it was for students to interact with GAPE, and how students ranked specific affordances of the technology in terms of its usability and usefulness. Students found GAPE relatively easy to use and appreciated its collaborative affordances. The researchers concluded that GAPE is a useful tool to meet learning objectives in the college composition classroom.

Editors’ Commentary

This paper explores the usability and effectiveness of Google Apps for Education, namely, to support writing (composition). The authors utilized a quantitative-based survey with college students, considering a wide range of factors such as ease of use, usefulness, and the desirability of various functionality. They found that students, despite many first-time users, rated the usability very high, including account set-up, login, text styling, and adding/reading comments. The participants also ranked features such as sharing documents, commenting, and collaboration as the most useful functionality. The authors recommended that educators consider the use of these type of apps for educating their students, particularly helping with writing skills. This article is a great example of exploring the usability and usefulness of a new technology, and how it might positively impact society.

#9: Macefield, R. (2009). [How to specify the participant group size for usability studies: A practitioner’s guide](#). *Journal of Usability Studies*, 5(1), 34–45.

Abstract

Using secondary literature, this article helps practitioners to specify the participant group size for a usability study. It is also designed to help practitioners understand and articulate the basis, risks, and implications associated with any specification. It is designed to be accessible to the typical practitioner and covers those study scenarios that are common in commercial contexts.

Editors' Commentary

A key part of UX research planning is to determine appropriate sample sizes. Macefield's literature review and synthesis provides guidance for practitioners, summarizing important mathematical concepts in problem discovery and statistical null hypothesis testing while avoiding the presentation of formulas that might push practitioners away. His primary takeaway from the literature is as important today as it was then—there are no magic numbers, no one-size-fits-all. Sometimes five is enough; sometimes it's more than you need, and sometimes you need more. The demands of usability studies focused on problem discovery (formative) are different from those focused on measuring differences between two products (summative). For example, larger samples in problem-discovery studies may be required when it is critical to detect low-probability issues in high-stakes or complex contexts. We expect this article to continue to be a valuable resource for usability and UX practitioners, but in cases where there is a need for more precise sample size estimates due to a high cost per participant, practitioners should consult with a statistician when planning their research.

#8: Sauro, J., & Zarolia, P. (2017). [SUPR-Qm: A questionnaire to measure the mobile app user experience](#). *Journal of Usability Studies*, 13(1), 17–37.

Abstract

In this paper, we present the SUPR-Qm, a 16-item instrument that assesses a user's experience of a mobile application. Rasch analysis was used to assess the psychometric properties of items collected from four independent surveys (N = 1,046) with ratings on 174 unique apps. For the final instrument, estimates of internal consistency reliability were high (alpha = .94), convergent validity was also high, with significant correlations with the SUPR-Q (.71), UMUX-Lite (.74), and likelihood-to-recommend (LTR) scores (.74). Scores on the SUPR-Qm correlated with the number of app reviews in the Google Play Store and Apple's App Store (r = .38) establishing adequate predictive validity. The SUPR-Qm, along with category specific questions, can be used to benchmark the user experience of mobile applications.

Editors' Commentary

This paper described an extension of Sauro's (2015) Standardized User Experience Percentile Rank Questionnaire (SUPR-Q; see #2 below) from a focus on website experiences to one on mobile app experiences. As far as we know, it is the only standardized usability or UX questionnaire to be developed using [item response theory](#) (IRT) instead of [classical test theory](#) (CTT), despite this having been initially proposed in the late 1990s (Hollemaans, 1999). While this background is of interest to UX researchers who develop standardized questionnaires, the paper's popularity is more likely due to practitioners' needs for standardized UX measures they can use in critical consumer spaces like mobile apps.

#7: Frederick, D., Mohler, J., Vorvoreanu, M., & Glotzbach, R. (2015). [The effects of parallax scrolling on user experience in web design](#). *Journal of Usability Studies*, 10(2), 87–95.

Abstract

Parallax scrolling is becoming an increasingly popular strategy in web design. This scrolling technique creates the illusion of depth on a webpage by making the background images move slower than the foreground images. In addition to its ability to engage users with a website, advocates of parallax scrolling claim that it improves user experience. Researchers attribute this pleasurable user experience to the fulfillment of the following variables: usability, satisfaction, enjoyment, fun, and visual appeal. We hypothesized that parallax scrolling would positively influence each of these five variables and subsequently the overall user experience. Eighty-six individuals from a large Midwestern university participated in the research. One group of participants (N = 43) interacted with the parallax scrolling website, whereas the second group (N = 43) interacted with the non-parallax scrolling website. An independent samples t-test revealed significant differences between the two groups in regards to perceived fun. Participants believed that the parallax scrolling website was more fun than the non-parallax scrolling website. The results of the study also showed parallax scrolling to be more effective when used

in a hedonic and fun context. Despite these benefits two of the participants suffered motion sickness and experienced significant usability issues while interacting with the parallax scrolling website. As a result, this potential risk to participants raises some ethical issues that UX practitioners and web designers should consider when planning to implement parallax scrolling.

Editors' Commentary

In this article, the authors focus on the user experience around parallax scrolling. This is an excellent example of an article that has a singular focus with an actionable outcome. In their research, they found that while there was no difference in satisfaction, there was a significant increase in perceived fun with parallax scrolling. In addition, they also cited a couple of participants who suffered motion sickness with parallax scrolling. One of the most appealing aspects of this paper is the practical aspect of the research. For example, the authors recommend that UX designers need to consider loading time when using parallax scrolling. Also, UX designers may consider the use of parallax scrolling to enhance the hedonic "fun" aspect of the experience.

#6: Wood, J., & Wood, L. (2008). [Card sorting: Current practices and beyond](#). *Journal of Usability Studies*, 4(1), 1–6.

Abstract

Card sorting was originally developed by psychologists as a method to the study of how people organize and categorize their knowledge. As the name implies, the method originally consisted of researchers writing labels representing concepts (either abstract or concrete) on cards, and then asking participants to sort (categorize) the cards into piles that were similar in some way. After sorting the cards into piles, the participants were then asked to give the piles a name or phrase that would indicate what the concepts in a particular pile had in common.

In the world of information technology, information architects and developers of desktop and Web-based software applications are faced with the problem of organizing information items, features, and functions to make it easier for users to find them. Card sorting can be an effective means of discovering the optimal organization of information for potential users' viewpoint.

Unfortunately, the development and practice of card sorting in information technology has been driven mostly by opinion and anecdotal experience, with little influence from systematic research. As a result, many of the decisions faced by developers in conducting a card sort study are based on analogy to seemingly similar arenas (e.g., survey administration) and situational circumstances. As we point out in the following sections, this assumption is often questionable.

Editors' Commentary

This essay does a magnificent job of summarizing the state of card sorting research. The authors utilize a simple format by summarizing a current practice, followed by a specific recommendation. For example, they offer a discussion and recommendations for the number of participants, participant instructions, understanding items, open vs. closed card sorting, and data analysis. They also highlight some areas in card sorting that need additional research, such as finding versus sorting, sorting time, and multi-level sorts. This essay is a must-read for anyone who is new to card sorting, looking for specific answers around card sorting, or might be curious about new research directions.

#5: Brooke, J. (2013). [SUS: A retrospective](#). *Journal of Usability Studies*, 8(2), 29–40.

Abstract

Rather more than 25 years ago, as part of a usability engineering program, I developed a questionnaire—the System Usability Scale (SUS)—that could be used to take a quick measurement of how people perceived the usability of computer systems on which they were working. This proved to be an extremely simple and reliable tool for use when doing usability evaluations, and I decided, with the blessing of engineering management at Digital Equipment Co. Ltd (DEC; where I developed SUS), that it was probably something that could be used by other organizations (the benefit for us being that if they did use it, we potentially had something

we could use to compare their systems against ours). So, in 1986, I made SUS freely available to a number of colleagues, with permission to pass it on to anybody else who might find it useful, and over the next few years occasionally heard of evaluations of systems where researchers and usability engineers had used it with some success.

Eventually, about a decade after I first created it, I contributed a chapter describing SUS to a book on usability engineering in industry (Brooke, 1996). Since then its use has increased exponentially. It has now been cited in (at the time of writing) more than 1,200 publications and has probably been used in many more evaluations that have not been published. It has been incorporated into commercial usability evaluation toolkits such as Morae, and I have recently seen several publications refer to it as an “industry standard”—although it has never been through any formal standardization process.

What is SUS, how did it get developed, and how has it reached this status of a de facto standard?

Editors’ Commentary

The System Usability Scale (SUS), despite humble beginnings, is a truly remarkable standardized questionnaire for the assessment of perceived usability (Lewis, 2018). In a review of 97 unpublished industrial usability studies, Sauro and Lewis (2009) found the SUS to be the most frequently used standardized usability questionnaire, accounting for 43% of post-test questionnaire usage. According to Google Scholar (as of 7/11/2021), the original SUS paper (Brooke, 1996) has been cited over 11,000 times. At JUS, we were fortunate for Brooke to contribute his retrospective of the origin and proliferation of the SUS, which itself has been cited over 1,500 times—a testimony to the importance of his contribution to the fields of usability and UX research and practice.

#4: Smelcer, J. B., Miller-Jacobs, H., & Kantrovich, L. (2009). [Usability of electronic medical records](#). *Journal of Usability Studies*, 4(2), 70–84.

Abstract

Health care costs represent a significant percentage of a country’s GDP. Implementing electronic medical records (EMR) systems are a popular solution to reducing costs, with the side benefit of providing better care. Unfortunately, 30% of EMR system implementations fail, often because physicians cannot use the EMRs efficiently. User experience problems, based on our experience at several clinics, are wide-spread among EMRs. These include loss of productivity and steep learning curves.

To help usability professionals contribute to the creation of more usable EMRs, we share our insights and experiences. Essential to understanding EMRs is the physician’s task flow, which we explain in detail. It is also helpful to understand the different work styles of physicians, variations in the pace of work, the use of nurses, the mode and timing of data entry, and variations in needed functionality. These variances in task flow, work styles, and needed functionality lead us to propose solutions to improve the usability of EMRs focusing on: flexible navigation, personalization and customization, accessing multiple patients, delegation of responsibility among medical personnel, and enabling data variations and visualizations.

Editors’ Commentary

This article addresses one of the most complex and critical design and experience spaces—electronic medical records. In this article, the authors explore some of the fundamental problems from the physician’s perspective, such as the long training times, loss of productivity, complexity of tasks, and complex functionality. They also adopt a different lens by examining the challenges to studying usability in clinics and hospitals. These include understanding terminology and privacy concerns, working with a wide range of different specialists, and of course understanding medical codes. To help with this, they share an in-depth tasks analysis. The article lays out a wonderful roadmap for future UX researchers in this space by suggesting design improvements to EMR’s by taking advantage of flexible navigation schemes, defaults, better handling of data variations, ability to access multiple patients, and more effective data visualization.

#3: Finstad, K. (2010). [Response interpolation and scale sensitivity: Evidence against 5-point scales](#). *Journal of Usability Studies*, 5(3), 104–110.

Abstract

A series of usability tests was run on two enterprise software applications, followed by verbal administration of the System Usability Scale. The original instrument with its 5-point Likert items was presented, as well as an alternate version modified with 7-point Likert items. Participants in the 5-point scale condition were more likely than those presented with the 7-point scale to interpolate, i.e., attempt a response between two discrete values presented to them. In an applied setting, this implied that electronic radio-button style survey tools using 5-point items might not be accurately measuring participant responses. This finding supported the conclusion that 7-point Likert items provide a more accurate measure of a participant's true evaluation and are more appropriate for electronically distributed and otherwise unsupervised usability questionnaires.

Editors' Commentary

A goal of UX research is to measure as accurately as possible. The literature on the effects of manipulating item formats is immense, including research in UX, psychology, social sciences, marketing, and politics. A continuing controversy is whether having more or fewer response options results in better measurement. Finstad contributed to this discussion using the unique method of presenting the SUS in a paper-and-pencil format to two groups, one marking the standard 5-point SUS and the other marking a 7-point version. He then checked to see how often respondents made a selection between response options, which he called an interpolation. The frequency of interpolations was not high, but all of them occurred in the 5-point condition. From a practical perspective, these interpolations would have little effect on resulting measurements, so, for example, we wouldn't recommend changing the SUS to an instrument made up of 7-point scales. On the other hand, when creating new questionnaires that have only one or a few items, these results support using at least 7-point instead of 5-point items, which Finstad did himself when developing the Usability Metric for Usability Experience (UMUX; Finstad, 2010).

#2: Sauro, J. (2015). [SUPR-Q: A comprehensive measure of the quality of the website user experience](#). *Journal of Usability Studies*, 10(2), 68–86.

Abstract

A three part study conducted over five years involving 4,000 user responses to experiences with over 100 websites was analyzed to generate an eight-item questionnaire of website quality—the Standardized User Experience Percentile Rank Questionnaire (SUPR-Q). The SUPR-Q contains four factors: usability, trust, appearance, and loyalty. The factor structure was replicated across three studies with data collected both during usability tests and retrospectively in surveys. There was evidence of convergent validity with existing questionnaires, including the System Usability Scale (SUS). The overall average score was shown to have high internal consistency reliability ($\alpha = .86$). An initial distribution of scores across the websites generated a database used to produce percentile ranks and make scores more meaningful to researchers and practitioners. The questionnaire can be used to generate reliable scores in benchmarking websites, and the normed scores can be used to understand how well a website scores relative to others in the database.

Editors' Commentary

It's one thing to use psychometric methods to create a standardized questionnaire, once that's done, the more difficult part is to develop norms that researchers and practitioners can use to interpret the resulting scores as good, bad, or indifferent. Sauro (2015) described a five-year effort to achieve that goal, publishing his original set of means and standard deviations with which researchers and practitioners could determine the percentile of measurements they made with the published SUPR-Q items and scales.

#1: Bangor, A., Kortum, P., & Miller, J. (2009). [Determining what individual SUS scores mean: Adding an adjective ratings scale](#). *Journal of Usability Studies*, 4(3), 114–123.

Abstract

The System Usability Scale (SUS) is an inexpensive, yet effective tool for assessing the usability of a product, including Web sites, cell phones, interactive voice response systems, TV applications, and more. It provides an easy-to-understand score from 0 (negative) to 100 (positive). While a 100-point scale is intuitive in many respects and allows for relative judgments, information describing how the numeric score translates into an absolute judgment of usability is not known. To help answer that question, a seven-point adjective-anchored Likert scale was added as an eleventh question to nearly 1,000 SUS surveys. Results show that the Likert scale scores correlate extremely well with the SUS scores ($r=0.822$). The addition of the adjective rating scale to the SUS may help practitioners interpret individual SUS scores and aid in explaining the results to non-human factors professionals.

Editors' Commentary

That brings us to our current #1 article—Bangor, Kortum, and Miller's paper describing a way to interpret SUS scores based on ratings they concurrently collected with an adjective rating scale. As mentioned above, standardization is a necessary step for a useful UX metric, but it is not sufficient. It can take even more work to develop methods for determining which standardized scores are good, mediocre, or bad. Bangor et al.'s original adjective rating scale had the following seven response options: Worst Imaginable, Awful, Poor, OK, Good, Excellent, and Best Imaginable. Current readers of this article should note that Bangor et al. expressed some reservation over the interpretation of "OK" (with an associated mean SUS of 50.9) as suggesting an acceptable experience given an overall mean SUS closer to 70 in their large-sample data (Bangor, Kortum, & Miller, 2008). Their current recommended practice is to anchor this response option with "Fair" instead of "OK" (Phil Kortum, personal communication, 22nd February 2018). Even without this adjustment, the adjective scale presented in this paper has increased the usefulness of the SUS to usability and UX practitioners, evidenced by its being our most popular article to date, having been cited almost 3,000 times.

Closing Thoughts

We hope you were able to discover some of the new research you may have missed, or perhaps refresh your memory of some articles you read many years ago. While we are very happy to share these ten articles with you, we also want to encourage you to explore our [entire inventory of articles](#) from our website. It is a rich, diverse, and useful catalog of some of the best user experience research in the world.

Acknowledgments

Many thanks to all who make the Journal of Usability Studies possible ... the UXPA (for paying the bills, providing guidance, and making this resource free to the UX community), the authors, the reviewers, and the production staff (Josh Rosenberg and Sarah Harris)!

References

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). [An empirical evaluation of the System Usability Scale](#). *International Journal of Human-Computer Interaction*, 24, 574–594.
- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. Taylor & Francis.
- Finstad, K. (2010). [The usability metric for user experience](#). *Interacting with Computers*, 22(5), 323–327.
- Holleman, G. (1999). [User satisfaction measurement methodologies: Extending the user satisfaction questionnaire](#). In Proceedings of *HCI International '99 8th International Conference on Human-Computer Interaction* (pp. 1008–1012), Munich, Germany. Lawrence Erlbaum Associates, Inc.

Lewis, J. R. (2018). [The System Usability Scale: Past, present, and future](#). *International Journal of Human-Computer Interaction*, 34(7), 577–590.

Sauro, J., & Lewis, J. R. (2009). [Correlations among prototypical usability metrics: Evidence for the construct of usability](#). In *Proceedings of CHI 2009* (pp. 1609-1618), Boston, MA. ACM.

About the Authors



Bill Albert, PhD

Dr. Albert is Executive Director of the User Experience Center and Adjunct Professor in Human Factors in Information Design at Bentley University. He is Co-Editor in Chief of the *Journal of Usability Studies*. Prior to joining Bentley, he was Director of User Experience at Fidelity Investments. He has published and presented his research at more than 30 national and international conferences. He co-wrote (with Tom Tullis) the first book on usability metrics, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics* (1st edition 2008, 2nd edition 2013). He also co-wrote (with Tom Tullis and Donna Tedesco) *Beyond the Usability Lab: Conducting Large-Scale Online User Experience Studies* in 2010. He received his PhD from Boston University. He completed a post-doc at Nissan Cambridge Basic Research.



James R. (Jim) Lewis, PhD, CHFP

Dr. Lewis is a Distinguished User Experience Researcher at MeasuringU. He has published influential papers on the assessment of perceived usability and sample size estimation for usability studies and has product experience in the design and evaluation of user interfaces (graphical, spoken, mobile), including having been the lead interaction designer for the product widely regarded as the first smart phone, the Simon. He is on over 90 US patents and has over 100 publications. His books include *Practical Speech User Interface Design* (2011), *Quantifying the User Experience* (with Jeff Sauro, 2nd ed., 2016), and *Using the PSSUQ and CSUQ in User Experience Research and Practice* (2019).