Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Effect of Perceived Ease of Use and Usefulness on UX and Behavioral Outcomes

James R. Lewis & Jeff Sauro

To cite this article: James R. Lewis & Jeff Sauro (24 Sep 2023): Effect of Perceived Ease of Use and Usefulness on UX and Behavioral Outcomes, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2260164

To link to this article: https://doi.org/10.1080/10447318.2023.2260164

Published online: 24 Sep 2023.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2260164

Taylor & Francis
Taylor & Francis Group

Effect of Perceived Ease of Use and Usefulness on UX and Behavioral Outcomes

James R. Lewis[a] and Jeff Sauro[b]

[a]MeasuringU, Delray Beach, FL, USA; [b]MeasuringU, Denver, CO, USA

ABSTRACT

We replicated and extended previous research to investigate the extent to which perceived ease of use and perceived usefulness account for variation in overall experience, likelihood to recommend, intention to use, and reported usage in a three-month follow-up. Consistent with previous research, we found little effect on structural equation models from varying three measures of perceived ease and two measures of perceived usefulness. All models had statistically significant standardized estimates and squared multiple correlations and had acceptable fit statistics. Despite these manipulations, the models supported a consistent narrative. Both perceived ease and perceived usefulness are important antecedents that either directly or indirectly affect the experiential and intentional outcomes (perceived usefulness somewhat more than perceived ease), with intention to use accounting for 19% of variation in follow-up ratings of usage. These models support UX practitioners by demonstrating the importance of work that improves perceptions of product ease and usefulness and showing that the two-item UX-Lite questionnaire is an effective and efficient measure of perceived ease and usefulness.

KEYWORDS

Standardized usability questionnaires; modified Technology Acceptance Model; mTAM; System Usability Scale; SUS; Usability Metric for User Experience; UMUX; UMUX-LITE; UX-Lite; likelihood-to-recommend; perceived usefulness; perceived ease of use; user experience; UX; behavioral intention; intention to use

1. Introduction

1.1. The importance of modeling UX drivers of experiential and intentional outcomes

In applied user experience (UX) research, there is a reasonable assumption that better user experiences lead to desirable business outcomes (e.g., increased likelihood to continue using or to recommend). Although these connections seem logical, UX researchers and practitioners would be in a stronger position to advocate for the importance of their work if there were models that quantified the extent to which key UX metrics such as perceived ease of use (PEoU) and perceived usefulness (PU) accounted for variation in outcome metrics such as ratings of overall experience, intention to use, and intention to recommend. Such models would also provide information about the relative importance of PEoU and PU as drivers of experiential and intentional outcomes.

1.2. Perceived usability, perceived ease of use, and perceived usefulness

The construct of perceived usability is most associated with the classical conception of usability (Brooke, 2013; ISO, 1998; Lewis et al., 2013, 2015; Sauro & Lewis, 2016), which is itself an important component of UX (Diefenbach et al., 2014; Lewis & Sauro, 2021). Sauro and Lewis (2009) demonstrated significant relationships between perceived usability and objective usability metrics (e.g., task completion times, successful task completion rates, and errors).

The constructs of PEoU and PU are most associated with the Technology Acceptance Model (TAM; Davis, 1989). According to TAM, the primary factors that affect a user's intention to use a technology are its perceived usefulness (PU) and perceived ease of use (PEoU). This model addressed early criticism of other models focusing only on product usability without assessment of usefulness (Pearson & Bailey, 1980). A number of studies support the validity of the TAM and its satisfactory explanation of end-user system usage (Wu et al., 2007). The TAM has undergone some evolution since its inception, but in the original version there were six items for each of its components.

The foundational paper (Davis, 1989) showed a correlation between the TAM and higher self-reported current usage ($r = 0.56$ for usefulness and $r = 0.32$ for ease of use), which is a form of concurrent validity. Participants were also asked to predict their future usage. This prediction had a strong correlation with ease and usefulness in the two pilot studies ($r = 0.85$ for usefulness and $r = 0.59$ for ease). But these correlations were derived from the same participants at the same time (not a longitudinal evaluation) which has the effect of inflating the correlation (people say they will use things more when they rate them higher).

In a longitudinal study by Davis et al. (1989), 107 MBA students were introducted to a word processor and then answered four usefulness and four ease of use items (subsets of the original TAM). Fourteen weeks later the same students completed the TAM again and answered self-reported usage questions. There was a strong correlation ($r = 0.63$) between behavioral intention and reported usage measured

CONTACT James R. Lewis jim@measuringu.com MeasuringU, Delray Beach, FL, USA

during the second phase ($r = 0.63$, not longitudinal), and a modest correlation between behavioral intention collected in the first phase and reported usage in the second phase ($r = 0.35$, longitudinal, accounting for about 12% of variation). PEoU and PU were significant drivers of intention, more so for PU than PEoU. In a later longitudinal study, Venkatesh and Speier (1999) reported significant correlations (about 0.58, accounting for 34% of variation) between behaviorial intentions and actual behavior at two periods after initial data collection (six and 12 weeks).

1.3. Modifying TAM from measure of expectation to measure of experience (mTAM)

As originally written, the 12 TAM items measure the extent to which people expect to experience an as-yet-unused product (e.g., "Using [this product] in my job would enable me to accomplish tasks more quickly" and "Learning to operate [this product] would be easy for me") with seven response options ranging from Extremely Unlikely to Extremely Likely.

Lewis (2019a) slightly modified the wording of the TAM items to enable their use with people who have experience using a product. Examples of these modified TAM (mTAM) items are "Using this product in my job enables me to accomplish tasks more quickly than other products in its class" and "Learning to operate this product was easy for me" with seven response options ranging from Extremely Disagree to Extremely Agree. In practice, mTAM item ratings are manipulated to produce a measure that can range from 0 to 100.

1.4. mTAM and other measures of perceived usability and usefulness

Lah et al. (2020) used data from three surveys to investigate the relationships between mTAM and alternate metrics of perceived usability and usefulness as drivers of overall experience and likelihood to recommend (LTR; Reichheld, 2003). The drivers (independent variables) they evaluated were the System Usability Scale (SUS; Brooke, 1996; Lewis, 2018b) and various metrics derived from the Usability Metric for User Experience (UMUX; Finstad, 2010; Sauro & Lewis, 2016).

The SUS is a popular measure of perceived usability. According to Google Scholar, Brooke (1996) has been cited almost 16,000 times in published research and the SUS accounts for an estimated 43% of post-study usage in unpublished usability studies (Sauro & Lewis, 2009). The SUS is a standardized questionnaire with 10 agreement items that have alternating tone, five response options from Strongly Disagree to Strongly Agree, and produces scores from 0 to 100 (Lewis, 2018b).

The UMUX (Finstad, 2010, 2013) is a standardized questionnaire with four agreement items that have alternating tone, seven response options from Strongly Disagree to Strongly Agree, and produces scores from 0 to 100. UX metrics derived from the UMUX include the UMUX-LITE (Lewis et al., 2013, 2015) and the UX-Lite® (Lewis & Sauro, 2021, Oct 12).

Both the UMUX-LITE and the UX-Lite are based on the two positive-tone items of the UMUX which provide alternate single-item measures of PEoU and PU (respectively, "[This product] is easy to use" and "[This product]'s capabilities meet my requirements") and, after interpolation, both produce scores from 0 to 100. They differ primarily in their number of response options, seven for the UMUX-LITE and five for the UX-Lite—a difference that does not appear to have much effect on their measurement properties (Lewis, 2021), but makes it easier to integrate the UX-Lite with other questionnaires that use 5-point items (e.g., SUS).

A potential advantage of these two-item measures of perceived ease and usefulness relative to standardized measures with more items (e.g., SUS, mTAM), is reduced user effort, especially if completing the questionnaire on a mobile device. Their smaller footprint also leaves room for other items in what might otherwise be overly lengthy surveys (as long as the reduction in the number of items doesn't substantially degrade their measurement properties).

The outcome metrics (dependent variables) in the Lah et al. (2020) models were a measure of overall experience ("Considering everything, how would you rate your overall experience with this product?" with 11 scale steps from Terrible to Excellent) and LTR ("Considering everything, how likely are you to recommend this product to a friend or colleague?" with 11 scale steps from Not At All Likely to Extremely Likely).

1.5. Key findings from Lah et al. (2020)

In three surveys conducted by Lah et al. (2020), respondents used SUS, UMUX-LITE and mTAM to rate their actual (as opposed to expected) experience with three different software products. As expected, the correlations between the mTAM measure of PEoU and the other measures of perceived usability (SUS and the UMUX-LITE Ease item) tended to be significantly stronger than correlations with measures of perceived usefulness (the mTAM measure of PU and the UMUX-LITE Usefulness item), evidence of construct validity for the distinction between PEoU and PU.

Also, a series of multiple regressions modeling the effects of PEoU and PU on the two outcome metrics (overall experience and LTR) were all statistically significant with reasonably consistent results across all three surveys. Substituting the SUS for mTAM PEoU had little effect on coefficients of determination or beta weights in the models, demonstrating that despite their historical and item content differences they appear to measure the same underlying construct (perceived usability).

Regression models using the UMUX-LITE Ease and Usefulness ratings in place of the mTAM PEoU and PU also had statistically significant coefficients of determination and beta weights. Relative to the mTAM models, the coefficients of determination in the UMUX-LITE models were slightly reduced but still substantial (averaging across models, from 66% to 60%). The consistency of outcomes for these regression models provides support for the use of the UMUX-LITE as a concise UX metric with theoretical and empirical connections to the TAM—effectively a mini-TAM.

1.6. Research goals

Continuing the investigation into the relationships among various measures of perceived usability (Lah et al., 2020; Lewis, 2018a, 2019b), the major goals of the current paper were to replicate and extend the Lah et al. models with a new dataset that had some variation in the independent (drivers) and dependent (outcome) variables.

To accomplish these goals, we investigated structural equation models (SEM) of the relationships between various measures of perceived usability (mTAM PEoU, SUS, UX-Lite Ease) and usefulness (mTAM PU, UX-Lite Usefulness) with outcome ratings of overall experience, LTR, intention to use, and reports of future use. Our expectations (based on the results reported by Lah et al., 2020) were:

- Models using the mTAM measures of PEoU and PU would be statistically significant with significant beta weights for PEoU and PU.
- Substituting SUS for mTAM PEoU would have little effect on magnitudes of coefficients of variation or beta weights.
- Substituting UX-Lite measures of Ease and Usefulness for mTAM PEoU and PU would result in statistically significant models with similar beta weights and some reduction in coefficients of variation.

2. Method

3.1. The participants

The participants were members of an online consumer panel, all from the United States. The percentages of males and females were about equal, with 66% below the age of 35. Respondents volunteered to participate in this research and were paid for participation by the online consumer panel. Participants cannot be identifed from their survey responses and, as is typical in consumer surveys, there was no risk associated with participation. We complied with the ethical standards of the Human Factors and Ergonomics Society (HFES, July 15, 2020) and the User Experience Professionals Association (UXPA, n.d.).

3.2. The surveys

Roughly every two years, surveys are conducted by MeasuringU to measure SUS and UX-Lite for about 60 software products (e.g., PowerPoint, Salesforce) along with ratings of overall experience and LTR. For the 2020 survey (designed and distributed using the MUIQ® framework), we also collected the mTAM and a three-item behavioral intention (BI) measure made up of the average of two items from TAM research (Venkatesh & Davis, 2000; "Assuming I had access to [Product], I intend to use it."; "Given that I had access to [Product], I predict that I would use it.") and a similar third item that we routinely collect ("I plan to use [Product] in the next three months"). The three BI items had seven response options from Strongly Disagree to Strongly Agree. At the beginning of the survey participants indicated which products they have used in the past year and were randomly assigned one of those to evaluate. We received complete sets of responses from 2,412 participants.

About three months after the initial survey, we followed up with a subset of respondents ($n = 321$) to find out how often they reported using their assigned product on a frequency scale with six response options: Never, Once a month, Once a week, Several times a week, Daily, Multiple times a day.

4. Results

Unless otherwise specified, statistical analyses used SPSS Version 23 (including AMOS Version 23 for structural equation modeling).

4.1. Survey 1: Ratings of business and consumer software products

4.1.1. Reliability

All of the questionnaires had values of coefficient alpha consistent with the prior literature. A common criterion for acceptable reliability is coefficient alpha equal to or greater than 0.70 (Nunnally, 1978). The values of coefficient alpha computed for the questionnaires were:

- SUS: 0.89.
- UX-Lite: 0.70.
- mTAM: 0.96 (with 0.95 and 0.95 respectively for PU and PEoU).
- BI: 0.94.

4.1.2. Concurrent validity

A common minimum criterion for evidence of concurrent validity is correlation greater than 0.30 between metrics (Nunnally, 1978). The correlations between SUS, UX-Lite (combined and by component), mTAM (combined and by component), overall experience, LTR, and BI ranged from 0.468 to 0.800 (all $p < 0.01$).

Consistent with the findings of Lah et al. (2020), the SUS had stronger correlations with the mTAM and UX-Lite measures of perceived ease than with their measures of perceived usefulness (non-overlapping 95% confidence intervals):

- SUS with mTAM PEoU: 0.800 (95% CI: 0.781–0.818).
- SUS with mTAM PU: 0.551 (95% CI: 0.517–0.581).
- SUS with UX-Lite Ease: 0.786 (95% CI: 0.764–0.806).
- SUS with UX-Lite Usefulness: 0.604 (95% CI: 0.571–0.635).

The pattern of correlations between mTAM and UX-Lite ease and usefulness components was also consistent with those of Lah et al. (2020). The correlation between mTAM PEoU and UX-Lite Ease was 0.763 (95% CI: 0.735–0.788), significantly higher than the correlation between mTAM PEoU and UX-Lite Usefulness of 0.675 (95% CI: 0.643–0.705). The correlation between mTAM PU and UX-Lite Usefulness was 0.697 (95% CI: 0.670–0.723), significantly higher than the correlation between mTAM PU and UX-Lite Ease of 0.493 (95% CI: 0.455–0.530).

4.1.3. Structural equation models

Figure 1 shows three structural equation models created with AMOS. The first one (Model A) used the components of the mTAM as drivers of overall experience, LTR, and BI. In Model B, mTAM PEoU was replaced with the SUS. In Model C, the mTAM PEoU and PU components were replaced with the UX-Lite Ease and Usefulness components.

The values on double-headed arrows are correlations between the primary drivers, values on single-headed arrows (links) are standardized estimates of the strengths of relationships between variables (interpreted like beta weights in multiple regression), and values above the upper right hand corners of outcome metrics are squared multiple correlations (interpreted like coefficients of determination in multiple regression—i.e., percentage of variance accounted for, also designated as R^2).

For example, in Model A, the correlation between mTAM PEoU and mTAM PU is 0.72, the strength of the connection between mTAM PEoU and BI (to use) is just 0.11 but between mTAM PEoU and overall experience (OverExp) is 0.42, and the percentage of variation in OverExp accounted for in the model is 59%. All correlations, standardized estimates, and squared multiple correlations in the models were statistically significant ($p < 0.0001$).

For assessing the goodness of fit of these types of models, we followed the advice of Jackson et al. (2009) who recommended reporting fit statistics that have different measurement properties such as the comparative fit index (CFI: a score of 0.90 or higher indicates good fit), the root-mean-square error of approximation (RMSEA: values less than 0.08 indicate acceptable fit), and the Bayesian information criterion (BIC: lower values are preferred). As shown in Figure 1, all three models had acceptable fit statistics, with Model C (UX-Lite drivers) nominally the best. Consistent with our expectation based on the results of Lah et al. (2020), the squared multiple correlations in Model C were lower than those in the other two models, but in most cases only by one or two percentage points (about five percentage points lower for BI relative to Model A). There was some variation from model to model in the magnitudes of correlations, squared multiple correlations, and standardized estimates, but the relative patterns from model to model were generally consistent. For example, the strength of connection

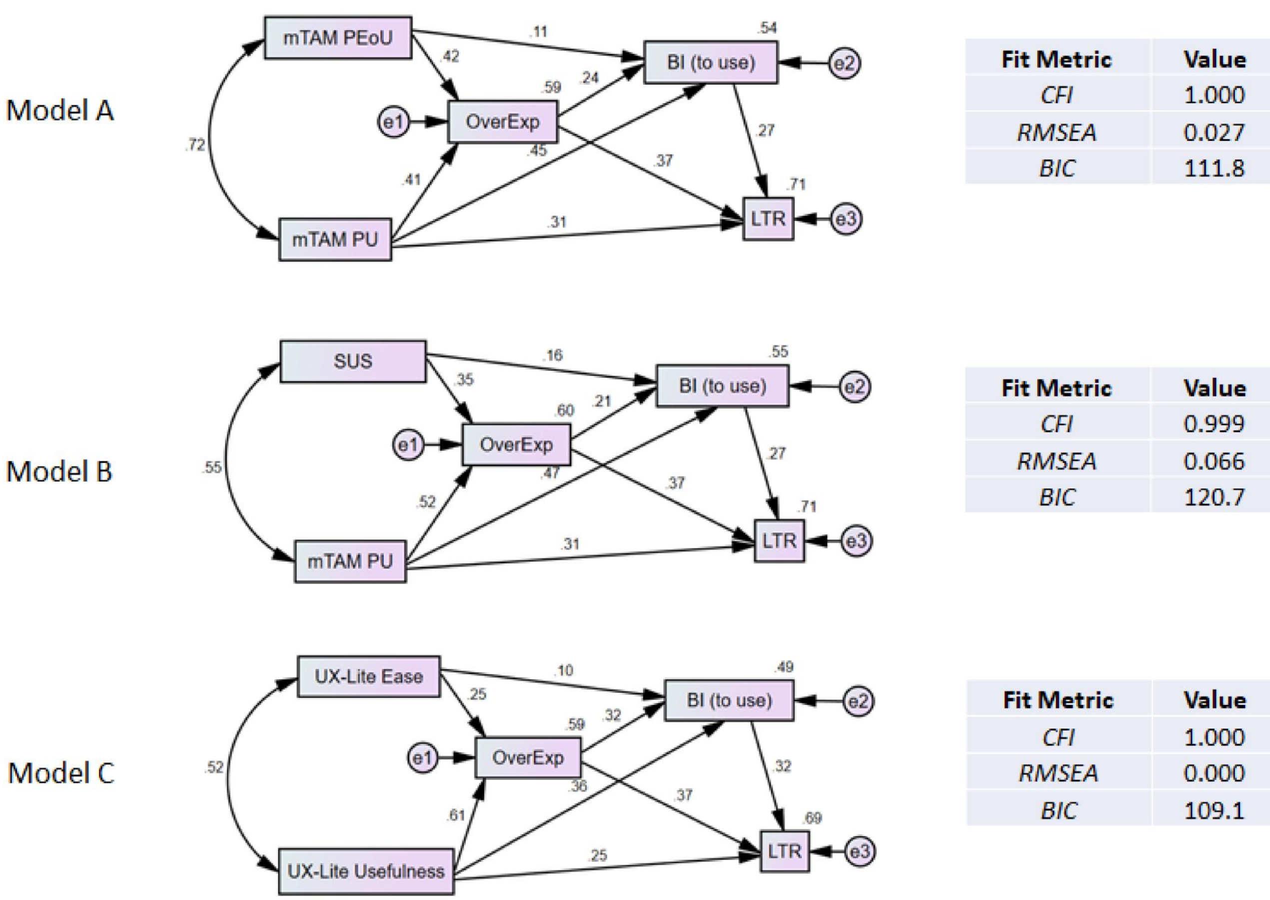


Figure 1. Structural equation models of perceived ease and perceived usefulness as drivers of overall experience, likelihood to recommend, and behavioral intention to use ($n = 2{,}412$).

between perceived ease/usability and the behavioral intention to use tended to be relatively weak, but the connection between perceived usefulness and the behavioral intention to use tended to be relatively strong.

An unexpected outcome was a significantly higher correlation between mTAM PEoU and mTAM PU (0.72, 95% confidence interval from 0.70 to 0.74) in Model A than for the corresponding predictors in Models B and C, respectively, a correlation of 0.55 (95% confidence interval from 0.52 to 0.58) between SUS and mTAM PU and a correlation of 0.52 (95% confidence interval from 0.49 to 0.55) between UX-Lite Ease and UX-Lite Usefulness.

4.2. Survey 2: Usage follow-up

Figure 2 shows the models from Figure 1 with the addition of the follow-up usage item collected from 321 participants in Survey 2. For the purpose of this analysis, the frequency scale for the follow-up item (how often the product rated in Survey 1 had been used over the past three months) had numeric values assigned to each response option (0: Never, 1: Once a month, 2: Once a week, 3: Several times a week, 4: Daily, 5: Multiple times a day).

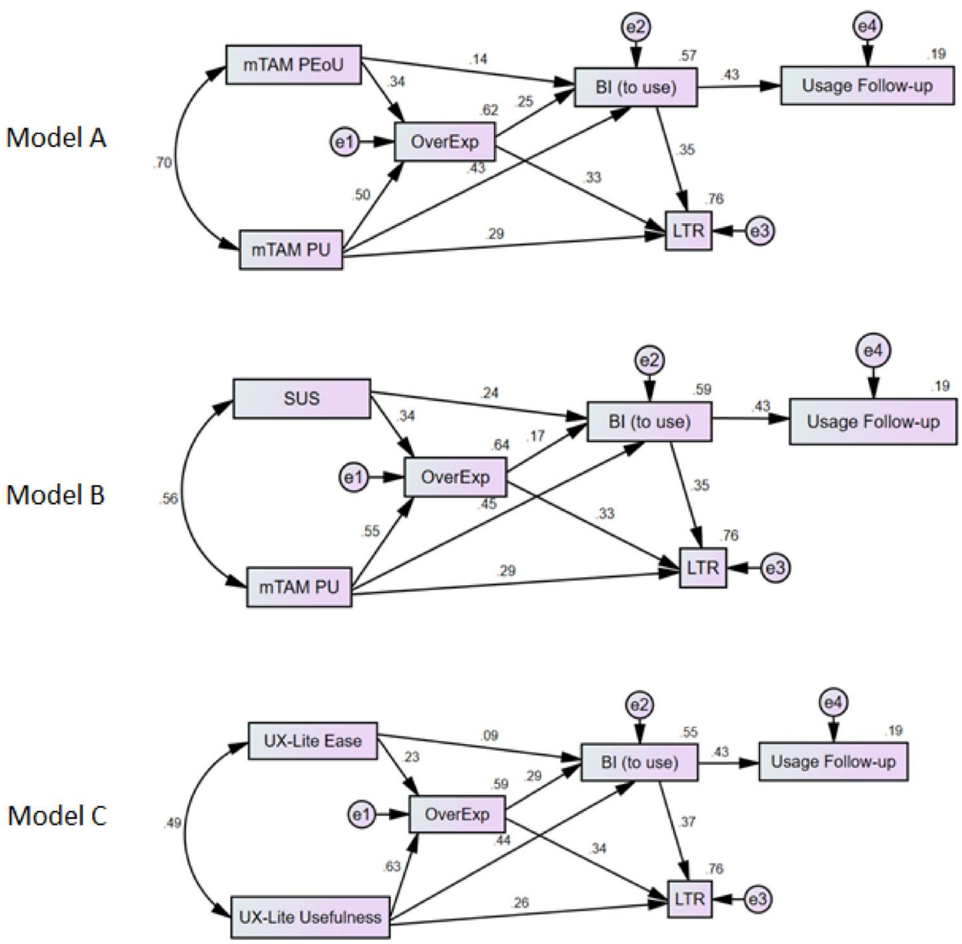
Even with the smaller sample size in this follow-up survey, all correlations (which had patterns similar to Survey 1 by being significantly larger for mTAM predictors in Model A than the predictors in Models B and C), standardized estimates, and squared multiple correlations in the models were statistically significant ($p < 0.01$), and all three models had similar (and acceptable) fit statistics. The models (specifically, behavioral intention) accounted for 19% of the variation in the usage follow-up ratings. The correlation between the primary predictors in Model A (0.70, 95% confidence interval from 0.64 to 0.75) was significantly higher than those in Models B and C (B: 0.56, 95% confidence interval from 0.48 to 0.63; C: 0.49, 95% confidence interval from 0.40 to 0.57).

5. Discussion

5.1. Psychometrics

All metrics used in the surveys (SUS, UX-Lite, and mTAM, including its PU and PEoU subscales) had acceptably high levels of reliability and acceptably high and statistically significant levels of concurrent validity (all $r > 0.45$, $p < 0.01$).

Analysis of convergent and divergent validity for metrics associated with the constructs of perceived ease and perceived usefulness replicated the findings of Lah et al. (2020), with ease metrics correlating significantly more



Fit Metric	Value
CFI	0.992
RMSEA	0.079
BIC	107.4

Fit Metric	Value
CFI	0.998
RMSEA	0.043
BIC	100.3

Fit Metric	Value
CFI	0.996
RMSEA	0.054
BIC	105.1

Figure 2. Structural equation models including measure of follow-up usage ($n = 321$).

with one another than with usefulness metrics (and vice-versa).

In summary, these results indicate that the measures used in these surveys had acceptable levels of the basic psychometric properties of reliability and validity.

5.2. Structural equation modeling

5.2.1. Statistical significance and goodness of fit

For all six SEMs in Figures 1 and 2, all correlations, standardized estimates, and squared multiple correlations were statistically significant ($p < 0.01$). All models had acceptable fit statistics.

5.2.2. Relative effects of perceived ease and usefulness

There was some variation in the standardized estimates for the effects of perceived ease and usefulness on overall experience. Both were significant drivers, but the effect of perceived usefulness usually tended to be greater. Perceived usefulness directly affected ratings of LTR and BI (to use), and indirectly affected them through its effect on overall experience. We did not find the link between perceived ease and LTR to be significant, so it was excluded from the model, and the standardized estimate between perceived ease and BI (to use) was relatively small. Perceived ease, however, had an indirect effect on LTR and BI (to use) through its effect on overall experience.

5.2.3. Substituting SUS for PEoU

The models were similar with regard to the magnitudes of standardized estimates and squared multiple correlations when substituting the SUS for PEoU. As Lewis (2018a) noted in a study of the correspondence of SUS, UMUX, UMUX-LITE, and the Computer System Usability Questionnaire (CSUQ), despite their historical and structural differences, all appeared to be measuring the same underlying construct, presumably, perceived usability. Consistent with the findings of Lah et al. (2020), these SEMs provide additional evidence that the PEoU component of the mTAM is a measure of the construct of perceived usability.

5.2.4. Substituting UX-Lite items for mTAM components

Again replicating the findings of Lah et al. (2020), substituting the Ease and Usefulness items of the UX-Lite for mTAM PEoU and PU produced reasonably consistent models, providing additional support for the use of the UX-Lite as a concise UX metric with theoretical and empirical connections to the TAM (a mini-TAM). In these models, there was less difference in the magnitudes of the squared multiple correlations for outcome variables than the differences in R^2 reported by Lah et al.

5.2.5. Usage follow-up

Adding the measure of usage follow-up to the models demonstrated that ratings of the behavioral intention to use accounted for 19% of variation in ratings of usage follow-up. They also illustrated the path from ratings of ease and usefulness through overall experience, LTR, and behavioral intention to use to evidence of actual use. Note that the estimate of 19% in this study is within the bounds of previous research, higher than the 12% reported by Davis et al. (1989) and lower than the 34% reported by Venkatesh and Speier (1999).

5.2.6. Correlations between various ease and usefulness predictors

Correlations between mTAM PEoU and mTAM PU (in Model A) were significantly higher than the corresponding predictors in Models B and C in both surveys, indicating less conceptual separation between the mTAM predictors than the others. Despite this difference, all three models across both surveys indicated similar structures with good fit indices.

5.3. Value of these models to UX practitioners

The original TAM research in the mid-1980s showed the effects of perceived ease and usefulness of products before use on the intention to use and, from there, actual use. This enabled information systems managers who wanted to maximize adoption of new products to focus on improving potential users' perceptions of the product's ease of use and usefulness without having to commit resources to the relatively difficult task of measuring outcomes (Davis et al., 1989).

UX industrial practitioners are often restricted to studying the extent to which design elements affect ratings of ease and usefulness, assuming that improving these aspects of user experiences will have positive effects on outcome metrics from behavioral intentions (e.g., intention to recommend, intention to use) to actual behaviors like recommending a product to others (a likely antecedent to additional sales) or continued use (a likely antecedent to additional subscriptions and reduced likelihood of defection). Borrowing the TAM strategy, these new models quantify the assumptions that perceived ease and usefulness drive key outcome measures, which confirms the value of doing the work to improve product ease and usefulness.

The variations in the models' measures of perceived ease and usefulness showed that mTAM PEoU, SUS, and the UX-Lite Ease item appear to be alternate measures of perceived usability. Similarly, the mTAM PU and UX-Lite Usefulness item appear to be alternate measures of perceived usefulness. This means that UX researchers and practitioners can use the simple two-item UX-Lite to measure perceived ease and usefulness in usability studies and UX surveys, saving time relative to the multi-item SUS and mTAM questionnaires.

5.4. Limitations and future research

As noted by Lah et al. (2020), much of the data in the current literature that examines the relationship among various measures of perceived usability (e.g., Lewis, 2018a, 2019b, Lewis et al., 2013, 2015), as well as the current research, are from surveys rather than usability studies. It would be good to replicate these findings with data from non-survey sources, but the barrier of collecting sufficiently large sample sizes from other sources will be difficult to overcome.

Despite our effort to collect data from multiple products, replication of this work with other user populations and products, especially if performed by other researchers, could enhance the generalizability of the findings.

6. Conclusions

We started by replicating the findings of Lah et al. (2020), which showed (1) a modified version of the TAM (mTAM) was predictive of overall experience and LTR and (2) similar model parameters when substituting SUS or UMUX-LITE Ease for mTAM PEoU or substituting UMUX-LITE Usefulness for mTAM PU. In the current research, we used SEMs that, in addition to perceived ease (measured with mTAM PEoU, SUS, or UX-Lite Ease), perceived usefulness (measured with mTAM or UX-Lite Usefulness), overall experience, and LTR, also modeled the outcome variable of behavioral intention to use and, in a second survey, added a measure of usage follow-up.

In all, we presented six SEMs, all of which had statistically significant standardized estimates and squared multiple corrrelations and acceptable fit statistics. The models differed in the specific metrics used to measure perceived ease and usefulness, and in the presence or absence of the measure of usage follow-up. Despite those manipulations, the models supported essentially the same narrative. Both perceived ease and perceived usefulness are important antecedents that either directly or indirectly affect overall experience, LTR, BI (to use), and reports of actual usage. Even though both are important, the effects of perceived usefulness seem to be more direct and somewhat stronger than the effects of perceived ease on the outcome metrics.

These models support UX practitioners by demonstrating the importance of work that improves perceptions of product ease and usefulness. They also show that UX researchers and practitioners can use the two-item UX-Lite in their work to effectively and efficiently measure perceived ease and usefulness.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

James R. Lewis http://orcid.org/0000-0002-3295-5392

References

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. Jordan, B. Thomas, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

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

Davis, D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–339. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Diefenbach, S., Kolb, N., Hassenzahl, M. (2014). The "Hedonic" in human-computer interaction: History, contributions, and future research directions. In *Proceedings of the 2014 Conference on Designing Interactive Systems – DIS 14* (pp. 305–314). Association for Computing Machinery.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Finstad, K. (2013). Response to commentaries on "The Usability Metric for User Experience". *Interacting with Computers*, *25*(4), 327–330. https://doi.org/10.1093/iwc/iwt005

HFES. (2020, July 15). *Code of ethics*. Human factors and ergonomics society. https://www.hfes.org/about-hfes/code-of-ethics.

ISO. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs), Part 11, Guidance on usability (ISO 9241-11: 1998E).*

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, P. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23. https://doi.org/10.1037/a0014694

Lah, U., Lewis, J. R., & Šumak, B. (2020). Perceived usability and the modified technology acceptance model. *International Journal of Human–Computer Interaction*, *36*(13), 1216–1230. https://doi.org/10.1080/10447318.2020.1727262

Lewis, J. R. (2018a). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human–Computer Interaction*, *34*(12), 1148–1156. https://doi.org/10.1080/10447318.2017.1418805

Lewis, J. R. (2018b). The System Usability Scale: Past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Lewis, J. R. (2019a). Comparison of four TAM item formats: Effect of response option labels and order. *Journal of Usability Studies*, *14*(4), 224–236.

Lewis, J. R. (2019b). Measuring perceived usability: SUS, UMUX, and CSUQ ratings for four everyday products. *International Journal of Human–Computer Interaction*, *35*(15), 1404–1419. https://doi.org/10.1080/10447318.2018.1533152

Lewis, J. R. (2021). Measuring user experience with 3, 5, 7, or 11 points: Does it matter? *Human Factors*, *63*(6), 999–1011. https://doi.org/10.1177/0018720819881312

Lewis, J. R., & Sauro, J. (2021). Usability and user experience: Design and evaluation. In G. Salvendy and W. Karwowski (Eds.), *Handbook of human factors and ergonomics* (pp. 972–1015). John Wiley.

Lewis, J. R., Sauro, J. (2021, October 12). *Evolution of the UX-Lite*. MeasuringU. https://measuringu.com/evolution-of-the-ux-lite/

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE – When there's no time for the SUS. In *Proceedings of CHI 2013.* (pp. 2099–2102). Association for Computing Machinery.

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring perceived usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction*, *31*(8), 496–505. https://doi.org/10.1080/10447318.2015.1064654

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Pearson, S. W., & Bailey, J. E. (1980). Measurement of computer user satisfaction. *ACM SIGMETRICS Performance Evaluation Review*, *9*(1), 59–68. https://doi.org/10.1145/1041872.1041881

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*(12), 46–54, 124.

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). Association for Computing Machinery.

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research* (2nd ed.). Morgan Kaufmann.

UXPA. (n.d.). *UXPA code of professional conduct*. User Experience Professionals Association. https://uxpa.org/uxpa-code-of-professional-conduct/

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., & Speier, C. (1999). Computer technology training in the workplace: A longitudinal investigation of the effect of mood. *Organizational Behavior and Human Decision Processes*, *79*(1), 1–28. https://doi.org/10.1006/obhd.1999.2837

Wu, J., Chen, Y., & Lin, L. (2007). Empirical evaluation of the revised end user computing acceptance model. *Computers in Human Behavior*, *23*(1), 162–174. https://doi.org/10.1016/j.chb.2004.04.003

About the authors

James R. Lewis is a distinguished user experience researcher at MeasuringU, an IBM Master Inventor emeritus (over 90 patents), the author (with Jeff Sauro) of Quantifying the User Experience, and a member of the Academy of Science, Engineering, and Medicine of Florida.

Jeff Sauro is the founder/CEO of MeasuringU. He is a pioneer in quantifying the user experience, widely recognized for making statistical concepts understandable and actionable, author of over 25 research papers and six books, including Benchmarking the User Experience, Customer Analytics for Dummies, and Quantifying the User Experience.