International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Perceived Usability and the Modified Technology Acceptance Model

Urška Lah, James R. Lewis & Boštjan Šumak

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2020.1727262

Perceived Usability and the Modified Technology Acceptance Model

Urška Lah[a,b], James R. Lewis [c], and Boštjan Šumak[b]

[a]EMO - Orodjarna Proizvodna družba, d. o. o, Celje, Slovenia; [b]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia; [c]MeasuringU, Denver, Colorado, USA

ABSTRACT

In response to recent criticism of the usefulness of the construct of usability, we investigated the relationships between measures of perceived usability and the components of a modified version of the Technology Acceptance Model (mTAM) – Perceived Usefulness (PU) and Perceived Ease-of-Use (PEU). In three surveys, respondents used SUS, UMUX-LITE and mTAM to rate their actual (as opposed to expected) experience with three software products. As expected, the correlations between PEU and other measures of perceived usability tended to be significantly stronger than those with PU. Additional findings support the use of the UMUX-LITE as a compact measure of perceived usability that has a strong relationship to the mTAM and strong correspondence with concurrently collected SUS scores. The main theoretical result of this research were regression results providing evidence that the PEU component of the mTAM appears to be another measure of the construct of perceived usability, connecting the TAM to the construct of perceived usability through the mTAM and providing evidence against the claim that the construct of usability is a theoretical dead end.

1. Introduction

1.1. Perceived usability and technology acceptance

In addition to the objective components of efficiency and effectiveness, perceived usability is an important component of the classical conception of usability (Brooke, 2013; ISO, 1998; Lewis, Utesch, & Maher, 2013, 2015; Sauro & Lewis, 2009, 2016), which is in turn a fundamental component of user experience (UX; Diefenbach, Kolb, & Hassenzahl, 2014). The first standardized usability questionnaires intended for application in usability testing appeared in the late 1980s (Brooke, 1996; Chin, Diehl, & Norman, 1988; Kirakowski & Dillon, 1988; Lewis, 1990).

Around the same time that usability researchers were producing the first standardized questionnaires to assess perceived usability, market researchers who studied the adoption of information systems were addressing similar issues. Of these, one of the most influential has been 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 (PEU). This model addressed early criticism of focusing only on usability without consideration of whether a product or system was useful (Pearson & Bailey, 1980).

A number of studies support the validity of the TAM and its satisfactory explanation of end-user system usage (Wu, Chen, & Lin, 2007). In the TAM, PU is the extent to which a person believes a technology will enhance job performance, and PEU is the extent to which a person believes that using the technology will be effortless. The more someone holds these beliefs before use, the greater their intention to use, and the more likely they are to try the technology. Figure 1 illustrates this model, and shows the wording of the items that Davis (1989) used to measure its constructs.

Recently, Tractinsky (2018), in a paper entitled "The Usability Construct: A Dead End?", argued against the usefulness of the construct of usability as a part of theory construction in human-computer interaction, in part due to "the inadequate modeling of the relations between the construct and its measures" (p. 133). In the same paper, he later cited the TAM as a good example of the use of constructs in scientific and practical model building, writing:

> Constructs contribute to a theory if they add to our understanding of the phenomenon under study. For example, the general domain that serves as the background for the emergence of the usability construct, namely, the use (often termed adoption or acceptance) of information technology, is a point of contact with various other theories. A notable such theory is the technology acceptance model (TAM; Davis, Bagozzi, & Warshaw, 1989). In TAM, the constructs "perceived ease-of-use" (a close relative of the intuitive meaning of "usability") and "perceived usefulness" are instrumental in explaining variations in the construct "behavioral intention" (people's intention to use a certain information system or product). Unfortunately, as previously mentioned, it is hard to name any influential theory in which the construct of "usability" plays a similarly useful role. (Tractinsky, 2018, p. 141)

The general reaction to the Tractinsky paper was that it offered valuable arguments regarding the difficulty of measuring usability and user experience, but those arguments were not universally accepted as the final word on the topic, especially with regard to the usefulness of usability as a construct

CONTACT James R. Lewis drjimnfl@gmail.com @drjimnfl; linkedin.com/in/jlewis6; MeasuringU, Denver, Colorado, USA

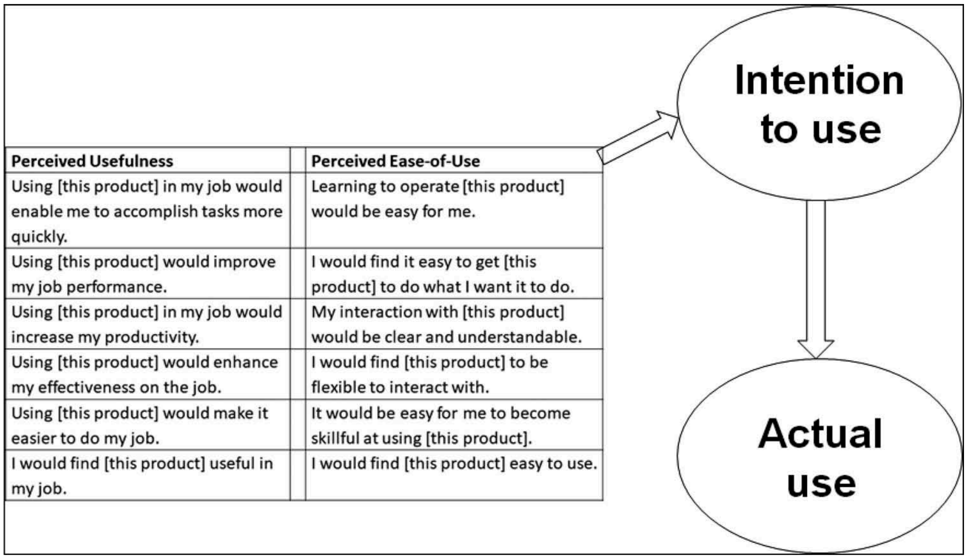

Figure 1. The Technology Acceptance Model.

guiding research and design. For example, Lewis (2018a) pointed out that discrepancies in reported correlations among typical usability measurements could be accounted for by differences in the scopes of literature reviews, and subscales of multidimensional questionnaires designed to assess different aspects of usability and user experience were typically correlated rather than uncorrelated, strongly suggesting the presence of a strong underlying and unifying factor presumed to be perceived usability. Borsci, Federici, Malizia, and De Filippis (2019) drew upon numerous responses to Tractinsky's paper to propose a constructive way to move forward with usability practice, calling upon the user experience community to implement mitigation strategies such as avoiding unnecessary fragmentation of knowledge while still permitting flexible application of usability/user experience standards via a meta-standard of usability and adoption of common guidelines to report usability data for cross-disciplinary communication.

Tractinsky's (2018) indictment of the usefulness of the construct of usability but endorsement of the construct of acceptance spurred us to investigate the statistical relationships among various measures of perceived usability and the components of the TAM. In particular, we were curious about the extent to which measures of perceived usability would correlate with measures of perceived usefulness and perceived ease-of-use. We were also interested in the extent to which measures of perceived usability and the components of the TAM were predictive of outcome metrics such as ratings of overall experience and likelihood-to-recommend (LTR). For measures of perceived usability, we focused on the System Usability Scale (Brooke, 1996; Lewis, 2018c) and various metrics derived from the Usability Metric for User Experience (Finstad, 2010; Sauro & Lewis, 2016).

1.2. The System Usability Scale (SUS)

With its roots in the 1980s, the SUS is a very widely used measure of perceived usability (Brooke, 1996, 2013; Lewis, 2018c), cited over 7300 times (according to Google Scholar) and accounting for an estimated 43% of post-study questionnaire usage in unpublished usability studies (Sauro & Lewis, 2009). For details regarding its use and scoring, see Lewis (2018c), and to review the specific version used in the current research, see Appendix Figure A1.

The SUS is popular for a number of reasons, including being in the public domain, having excellent psychometric properties, and, since 2008, having a substantial amount of normative research which has led to several ways to interpret the SUS (Bangor, Kortum, & Miller, 2008; 2009; Lewis, 2018c). Table 1 shows one way to interpret SUS scores, using a curved grading scale (CGS) developed by Sauro and Lewis (2016). The data for the CGS came from a collection of 446 studies (over 5000 individual SUS responses). The data were normalized with a logarithmic transformation on reflected scores, then organized in the table. For quantitative analyses based on grades, the table includes a column of standard grade points associated with the letter grades.

1.3. Usability Metric for User Experience (UMUX)

The UMUX (Finstad, 2010, 2013) is a relatively recent addition to the inventory of standardized usability questionnaires. For details regarding its use and scoring, see Lewis (2018b, 2019b), and to review the specific version used in the current

Table 1. The Sauro–Lewis curved grading scale with grade points.

SUS Score Range	Grade	Grade Point	Percentile Range
84.1 – 100	A+	4.0	96-100
80.8–84.0	A	4.0	90-95
78.9–80.7	A-	3.7	85-89
77.2–78.8	B+	3.3	80-84
74.1–77.1	B	3.0	70-79
72.6–74.0	B-	2.7	65-69
71.1–72.5	C+	2.3	60-64
65.0 – 71.0	C	2.0	41-59
62.7–64.9	C-	1.7	35-40
51.7–62.6	D	1.0	15-34
0.0–51.6	F	0.0	0-14

research, see Appendix Figure A2. A key goal of UMUX development was to obtain a metric that produced scores similar to the SUS but with fewer items (the SUS has 10 items, the UMUX has four).

In addition to the initial research by Finstad (2010), other researchers (Berkman & Karahoca, 2016; Borsci, Federici, Gnaldi, Bacci, & Bartolucci, 2015; Lewis, 2018b, 2019b; Lewis et al., 2013, 2015) have also reported desirable psychometric properties for the UMUX, including acceptable levels of reliability, concurrent validity and sensitivity to different levels of a variety of independent variables. Most research in this area has found substantial correspondence between the magnitudes of mean SUS and UMUX scores.

1.4. UMUX-LITE

The UMUX-LITE is a short version of the UMUX consisting of its positive-tone items (Items 1 and 3), selected on the basis of factor and item analysis (Lewis et al., 2013), which are:

- UMUX01: This system's capabilities meet my requirements.
- UMUX03: This system is easy to use.

In addition to the statistical analyses supporting their selection, it is interesting that the content of the two items has similarities to the constructs of the TAM (Davis, 1989), with the first item (meets requirements) being an aspect of usefulness and the second directly assessing ease-of-use.

Research on the UMUX-LITE and UMUX-LITEr (Berkman & Karahoca, 2016; Borsci et al., 2015; Lewis et al., 2013, 2015) has demonstrated acceptable psychometric properties, including:

- Acceptable reliability (coefficient alpha ranging from 0.77 to 0.86)
- Concurrent validity (correlations with SUS ranging from 0.74 to 0.83; correlation with ratings of likelihood-to-recommend ranging from 0.72 to 0.74)
- Sensitivity (significant differences as a function of respondents' ratings of frequency-of-use)

There are multiple versions of the UMUX-LITE to be aware of, and the UMUX-LITE literature has been inconsistent in its terminology (for details, see Lewis, 2019b). To summarize, two versions of UMUX-LITE have been published, one with a regression adjustment and one without. The preferred designation of the regression-adjusted version is UMUX-LITEr and for the unadjusted version, UMUX-LITE. The purpose of the regression adjustment was to bring UMUX-LITE scores into closer correspondence with concurrently collected SUS scores.

The correspondence in the magnitudes of UMUX-related metrics and the SUS has been investigated a number of times since its initial publication. Recently, Lewis (2019b) reported that range restriction associated with the UMUX-LITEr appeared to affect the correspondence of scores for products receiving above-average SUS scores on the CGS (see Table 1, specifically, when grades were B+ or better, SUS ≥ 77.2). This is not surprising in retrospect because the data used to develop the regression model were collected on products with average levels of perceived usability, apparently biasing the formula against accurate measurement of higher levels of perceived usability. In reaction to this, Lewis (2019b) suggested that practitioners use the unadjusted UMUX-LITE rather than UMUX-LITEr.

1.5. Technology Acceptance Model (TAM)

The TAM questionnaire is made up of 12 items, six for the measurement of PU and six for PEU, as shown in Figure 1. The original TAM rating scales were designed to elicit likelihood ratings rather than agreement ratings because the purpose of the model was to predict future use of a product (anticipated usefulness and ease-of-use as perceived before any use) rather than to rate the experience of actual use (perceived usability after use as opposed to objective measures of usability). Also, because the conceptual definition of PU was on the enhancement of job performance, most of the associated items directly reference a work context.

Davis (1989) conducted a lab study in which 40 participants evaluated (in counterbalanced order) two graphics applications with different user interfaces. Coefficient alpha was 0.98 for PU and 0.94 for PEU, and multitrait-multimethod analyses indicated appropriate convergent and divergent validity. Factor analysis of the data showed the expected pattern of association of items with factors. Both PU and PEU correlated significantly with self-predictions of likelihood of use if the product were available at the participants' place of work.

Because the purpose of the current research was to evaluate PU and PEU after product experience rather than predicting the likelihood of use, the TAM was slightly revised for our purposes to match the format of the UMUX. With this modification respondents can rate the strength of agreement with items regarding actual user experience rather than likelihood of future use. To avoid informing participants about the expected alignment of items with PU and PEU constructs, the 12 items were presented as a single questionnaire (the modified TAM, or mTAM – see Appendix Figure A3). To get mTAM scores that, like the SUS and UMUX, range from 0 to 100, for PU and PEU separately, compute the mean of the item scores, subtract one from that mean, then multiply by 100/6. To get an overall mTAM score, compute the mean of PU and PEU.

1.6. Experiential outcomes: Likelihood-to-recommend and overall experience

In the original TAM research, the focus was on the prediction of likelihood-of-use before experiencing the product. After experiencing a product, it is reasonable to shift to the prediction of outcomes such as likelihood-to-recommend and assessment of the overall experience.

Likelihood-to-recommend (LTR) is a popular measure of customer loyalty, probably best known for its use in the Net Promoter Score (Reichheld, 2003) and is most applicable when users have a choice of which product to use (Lewis, 2018c; Sauro & Lewis, 2016). In its most common form, the item is "Considering everything, how likely are you to recommend this product to a friend or colleague?" with 11 scale steps from 0 (Not at all likely) to 10 (Extremely likely). Regression analyses of concurrently collected SUS and LTR data from 2201 users and over 80 products found a strong positive correlation of .62, meaning SUS scores explained about 39% of the variability in responses to the LTR item (Sauro & Lewis, 2016). This leaves 61% of its variability unexplained, so although SUS and LTR have a strong relationship, there may be factors other than perceived usability that affect likelihood-to-recommend.

After experiencing a product, users can form an overall opinion of their experience. There are many ways to capture this attitude. One quick way is to collect ratings with an item modeled on likelihood-to-recommend, specifically, "Considering everything, how would you rate your overall experience with this product?" with 11 scale steps from 0 (Terrible) to 10 (Excellent). Research has consistently shown significant correlations between this measure of overall user experience with LTR and other measures of perceived usability (Lewis, 2018c).

1.7. Research goals

As part of a continuing investigation into the relationships among various measures of perceived usability (Lewis 2018b, 2019b), the major goals of the current paper were to extend this line of research to measures of technology acceptance using mTAM and to extend research on the correspondence of measurement between SUS and UMUX-LITE to additional products and an additional language, Slovene.

To accomplish these goals, we investigated the following across three surveys:

- The relationship between measures of perceived usability and concurrently collected measures of the PU and PEU components of the mTAM, with the expectation that measures of perceived usability would have a significantly higher correlation with PEU than with PU, in other words, that perceived ease-of-use should be more strongly related than perceived usefulness to measures of perceived usability.
- Regression models of the relationship between PU and PEU with outcome ratings of LTR and overall experience, with the expectations that (a) substituting standardized measures of perceived usability for PEU would not affect magnitudes of beta weights or coefficients of determination and (b) models using the items of the UMUX-LITE would be similar to models using the components of mTAM, demonstrating structural connections among the mTAM and standardized usability measures in the context of their relationships with key outcome metrics.
- The correspondence of concurrently collected UMUX, UMUX-LITE and UMUX-LITEr means with the mean SUS, with the expectation that the UMUX-LITE would provide better correspondence than the UMUX-LITEr over a wide range of measures of perceived usability.

2. Method

We conducted three surveys that, to enhance generalizability, collectively covered two different user populations, three different products, and two different languages. Unless otherwise specified, statistical analyses used SPSS Version 25. When referring to specific mTAM, SUS, or UMUX items, the convention in this paper is to provide a two-digit item number after the questionnaire's acronym, for example, SUS10 for the tenth item of the SUS, or mTAM06 for the sixth item of the mTAM.

2.1. Survey 1

Participants in Survey 1 rated their experience using PowerPoint. There were three versions of the survey which differed in the order of presentation of the usability questionnaires (Latin square design: mTAM/SUS/UMUX, SUS/UMUX/mTAM, UMUX/mTAM/SUS). After random assignment to a survey version, participants completed the three usability questionnaires and a section which included overall assessment items (overall experience and likelihood to recommend), system questions (type of hardware, operating system and types of applications typically used by the respondent) and usage characteristics (length of time using the rated product, and frequency of product use). In accordance with the rules established by their developers, missing data in the SUS and UMUX were replaced with the center item of the rating scale (3 for the SUS, 4 for the UMUX). For consistency, this rule was also followed for the mTAM (missing data replaced with center scale value of 4).

Respondents were members of the IBM User Experience panel. In early 2017, the initial panel was formed by emailing invitations to 20,000 randomly selected IBM US employees of whom about 10% agreed to join the panel. At the time of this study, there were 2155 members in the panel. As a check on data quality, cases in which respondents had more than a 50-point difference among the SUS, UMUX, or mTAM scores (after translation to a common 0–100 point scale) were removed based on the likelihood that those respondents had made rating errors due to the different questionnaire formats. The total number of respondents who completed the survey was 488, with 39 having one or two missing responses replaced as described above. Five cases were deleted as a result of the data quality check, leaving 483 for inclusion in the analyses.

2.2. Survey 2

Survey 2 was modeled after Survey 1, with the following differences. The rated product was Gmail. Respondents were members of different Slovenian educational faculties and employees of selected Slovenian companies, with invitations emailed in June 2018 to approximately 3000 potential participants. All Slovenian respondents had experience using Gmail, but because it was not necessarily the e-mail program used for their work, the Slovenian mTAM items were modified to remove references to "my/the job". The total number of respondents who sufficiently completed the Gmail survey was 412, with 22 (5.0%) having one or two missing responses replaced as described above. Fifteen cases (3.6%) were deleted as a result of the data quality check, leaving 397 for inclusion in the analyses.

A Slovene version of the SUS was available (Blažica & Lewis, 2015) but it was necessary to translate the items for the UMUX and mTAM. First, 10 reviewers from the field of computer sciences individually reviewed a draft translation, with their input incorporated into a revised draft. The second stage was to perform a back-translation in which a translation agency, without reference to the original, translated the final draft back into English. This draft was adjusted in the case of two words for UMUX, "neprijetna" ("unpleasant") adjusted to "me frustrira" ("frustrating") and "zmožnosti" ("capacities") adjusted to "zmogljivosti" ("capabilities"). Finally, two independent translators, without reference to the original, translated the adjusted final draft back into English. The translators were native Slovene speakers fluent in English. For all items, both translators provided back translations with the same meaning as the original. Table 2 shows the original English and final Slovene versions of the questionnaires.

2.3. Survey 3

Survey 3 was modeled after Survey 1, with the following differences. Participants (again IBM employees who were members of the User Experience panel) rated their experience using IBM Notes. There were four versions of the survey which differed in the version of the mTAM presented to participants (different item response formats and orders). Because there was no order effect in either Survey 1 or 2 (see below), all surveys presented the three questionnaires in the same order: mTAM/SUS/UMUX. The total number of participants who completed the survey was 568, with 38 having one or two missing responses replaced as described above. Twenty-two cases were deleted as a result of the data quality check, leaving 546 for inclusion in the analyses. Analyses of the different versions of mTAM used in this survey are available in Lewis (2019a). In summary, there were no significant differences as a function of format group in either comparisons of means or factor analyses, so the data were combined for the following analyses.

3. Results

3.1. Survey 1: Ratings of PowerPoint (United States)

3.1.1. Order of presentation of the questionnaires

Each participant experienced one of three orders of presentation of the questionnaires, with the order conforming to a Latin square design. There was no effect of the order of presentation ($F(2, 480) = 2.1$, $p = .13$ – less than a 1 point difference between the smallest and largest means), so the data were combined for the following analyses.

3.1.2. 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: 0.91
- UMUX: 0.85
- UMUX-LITE: 0.73
- mTAM: 0.95 (with 0.95 and 0.95, respectively, for the PU and PEU subscales)

3.1.3. Concurrent validity

A common minimum criterion for evidence of concurrent validity is a correlation between metrics that exceeds 0.3 (Nunnally, 1978). Table 3 shows the correlations among the various measures of perceived usability and usefulness in this survey. All correlations in Table 3 were statistically significant, but were not all of the same magnitude.

Of particular interest are the correlations (with 95% confidence intervals) between measures of perceived usability and the mTAM components of PU and PEU. Figure 2 shows clear discrimination between the correlations of these components, with PEU correlations significantly higher ($p < .05$) than correlations with PU.

Table 2. Translated mTAM and UMUX items.

English Version of mTAM (modified for non-work setting)	Slovene Version (mTAM-SI)
1. Using Gmail enabled me to accomplish tasks more quickly than my previous e-mail product.	1. Uporaba Gmaila mi omogoča hitrejše opravljanje nalog v primerjavi z drugimi e-poštnimi rešitvami.
2. Using Gmail improved my job performance.	2. Uporaba Gmaila je izboljšala mojo delovno uspešnost.
3. Using Gmail increased my productivity.	3. Uporaba Gmaila je zvišala mojo produktivnost.
4. Using Gmail enhanced my effectiveness.	4. Uporaba Gmaila je povečala mojo učinkovitost.
5. Using Gmail made it easier to do the things I needed to do with it.	5. Uporaba Gmaila je olajšala opravila, ki sem jih moral opraviti.
6. I found Gmail useful.	6. Gmail se mi je zdel uporaben.
7. It was easy to learn to operate Gmail.	7. Enostavno se je bilo naučiti uporabljati Gmail.
8. I found it easy to get Gmail to do what I wanted it to do.	8. Z Gmailom sem enostavno prišel do želenih rezultatov.
9. My interaction with Gmail was clear and understandable.	9. Interakcija z Gmailom je bila jasna in razumljiva.
10. I found Gmail to be flexible to interact with.	10. Interakcija z Gmailom se mi je zdela fleksibilna.
11. It was easy for me to become skillful at using Gmail.	11. Enostavno je bilo postati spreten pri uporabi Gmaila.
12. I found Gmail easy to use.	12. Gmail se mi je zdel enostaven za uporabo.
English Version of UMUX	**Slovene Version (UMUX-SI)**
1. Gmail's capabilities meet my requirements.	1. Zmogljivosti Gmaila izpolnjujejo moje zahteve.
2. Using Gmail is a frustrating experience.	2. Uporaba Gmaila me frustrira.
3. Gmail is easy to use.	3. Gmail je enostaven za uporabo.
4. I have to spend too much time correcting things with Gmail.	4. Preveč časa porabim za popravljanje napak ob uporabi Gmaila.
The anchors: strongly disagree 1 2 3 4 5 6 7 strongly agree	The anchors: sploh se ne strinjam 1 2 3 4 5 6 7 se povsem strinjam

Table 3. Correlations among various measures of perceived usability and usefulness for Survey 1.

	SUS								
UMUX	0.855	**UMUX**							
UMUX-LITE	0.816	0.897	**UMUX-LITE**						
UMUX01	0.637	0.739	0.870	**UMUX01**					
UMUX03	0.801	0.848	0.906	0.578	**UMUX03**				
mTAM	0.801	0.755	0.774	0.665	0.709	mTAM			
PU	0.613	0.586	0.621	0.600	0.511	0.904	**PU**		
PEU	0.835	0.778	0.778	0.601	0.770	0.903	0.633	**PEU**	
OverExp	0.802	0.791	0.778	0.688	0.695	0.826	0.704	0.788	**OverExp**
LTR	0.746	0.731	0.745	0.692	0.637	0.807	0.730	0.730	0.896

All correlations had at least 480 df and were statistically significant (p < 0.01).

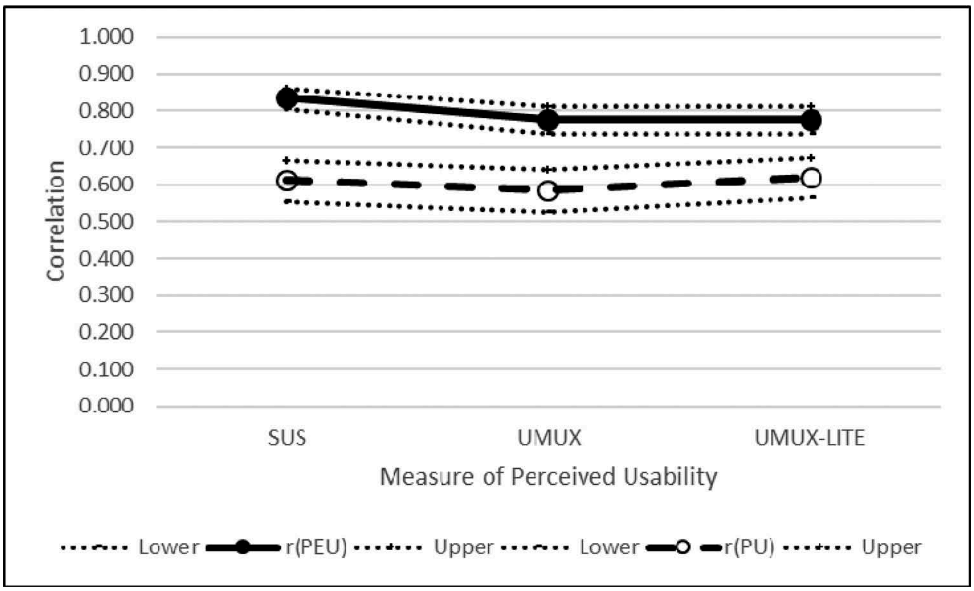

Figure 2. Correlations (with 95% confidence intervals) between perceived usability and mTAM components for Survey 1.

Also of interest were the differences in correlation between PU and PEU with UMUX01 (useful: functional adequacy) and UMUX03 (easy to use), the two items that make up the UMUX-LITE. The correlation between PU and UMUX01 (0.600) was significantly greater than PU with UMUX03 (0.511; $Z = 1.99$, $p = .047$). The reverse was the case for PEU, for which the correlation with UMUX01 (0.601) was significantly less than its correlation with UMUX03 (0.770; $Z = 5.03$, $p < .0001$).

Table 4. Adjusted R^2 and beta weights for regression analyses in Survey 1.

Predicting	R^2adj	Beta 1	Beta 2
LTR with PU and PEU	65%	0.446	0.446
LTR with UMUX01 and UMUX03	56%	0.486	0.355
LTR with PU and SUS	67%	0.436	0.477
Overall Experience with PU and PEU	69%	0.314	0.570
Overall Experience with UMUX01 and UMUX03	61%	0.429	0.448
Overall Experience with PU and SUS	72%	0.342	0.593

Beta weights 1 and 2 are those for the "predicting with" variables, in that order (e.g., for the second row, the beta weight for UMUX01 was 0.486 and for UMUX02 was 0.355).

3.1.4. Construct validity

Parallel analysis indicated one-factor solutions for the SUS and UMUX, but two factors for the mTAM. Factor analysis of the mTAM (unweighted least squares with Varimax rotation) produced the expected alignment of items with factors. Items 1–6 aligned on the same factor with weights ranging from 0.683 to 0.891; Items 7–12 aligned on the other factor with weights ranging from 0.743 to 0.861.

3.1.5. Regression analyses

Table 4 shows the adjusted coefficient of determination (R^2adj) and beta weights for regression analyses (1) using mTAM and UMUX-LITE to predict LTR and overall experience ratings and (2) substituting SUS for PEU. The coefficient of determination indicates the percentage of variation in the dependent (predicted) variable that is accounted for by the independent (predictor) variables, so larger values indicate a stronger model. Beta weights are standardized regression coefficients for which larger values indicate stronger effects on prediction. All values in Table 4 were statistically significant ($p < .05$).

As shown in the first row of Table 4, PU and PEU accounted for 65% of the variation in LTR (95% confidence interval ranging from 60-70%) with significant beta weights for both components. Replacing PU and PEU with the two items of the UMUX-LITE, as shown in the second row, produced a similar outcome, accounting for 56% of the variation in LTR (95% confidence interval ranging from 50-62%).

Although there appeared to be some reduction in the coefficient of determination when using the UMUX-LITE items in place of the mTAM components, overlap in confidence intervals indicated the difference was not statistically significant. Substitution of SUS for PEU produced almost identical outcomes for the coefficient of determination and beta weights.

For prediction of overall experience, PU and PEU accounted for 69% of the variation in LTR (95% confidence interval ranging from 64-73%) with significant beta weights for both components. Replacing PU and PEU with the two items of the UMUX-LITE (fifth row), produced a similar outcome, accounting for 61% of the variation in LTR (95% confidence interval ranging from 55-66%). Again, the substitution of SUS for PEU in the model produced almost identical outcomes with regard to variation accounted for and magnitude of beta weights.

3.2. Survey 2: Ratings of Gmail (Slovenia)

3.2.1. Order of presentation of the questionnaires

Each participant experienced one of three orders of presentation of the questionnaires, with the order conforming to a Latin square design. The effect of order of presentation was not significant ($F(2, 394) = 2.66$, $p = .07$ – with a 1.8-point difference between the smallest and largest means), so the data were combined for the following analyses.

3.2.2. Reliability

All of the questionnaires had acceptable values of coefficient alpha:

- SUS: 0.88
- UMUX: 0.79
- UMUX-LITE: 0.69
- mTAM: 0.95 (with 0.93 and 0.95, respectively, for the PU and PEU subscales)

3.2.3. Concurrent validity

Table 5 shows the correlations among the various measures of perceived usability and usefulness in this survey. As in Survey 1, all correlations in Table 5 were statistically significant, but were not all of the same magnitude. Figure 3 shows the same clear discrimination between the correlations of these components as in Figure 2, with PEU correlations significantly higher ($p < .05$) than correlations with PU. Unlike Survey 1, the correlation between PU and UMUX01 (0.39) was not significantly greater than PU with UMUX03 (0.47; $Z = 1.4$, $p = .17$). For PEU the correlation with UMUX01 (0.58) was significantly less than its correlation with UMUX03 (0.72; $Z = 3.4$, $p = .0006$).

3.2.4. Construct validity

The parallel analyses were the same as Survey 1, indicating one-factor solutions for SUS and UMUX but a two-factor solution for mTAM. For the factor analysis of the mTAM (unweighted least squares with Varimax rotation), the alignment of items with mTAM factors was almost, but not quite, as expected because Item 6 aligned with Items 7–12 (with factor loadings ranging from 0.661 to 0.894) instead of with Items 1–5 (factor loadings ranged from 0.600 to 0.933).

Table 5. Correlations among various measures of perceived usability and usefulness for Survey 2.

	SUS								
UMUX	0.780	**UMUX**							
UMUX-LITE	0.740	0.874	**UMUX-LITE**						
UMUX01	0.567	0.751	0.877	**UMUX01**					
UMUX03	0.729	0.777	0.871	0.527	**UMUX03**				
mTAM	0.703	0.626	0.669	0.529	0.641	**mTAM**			
PU	0.521	0.453	0.493	0.395	0.467	0.925	**PU**		
PEU	0.780	0.707	0.746	0.583	0.721	0.887	0.645	**PEU**	
OverExp	0.644	0.645	0.664	0.562	0.599	0.671	0.576	0.648	**OverExp**
LTR	0.595	0.611	0.618	0.526	0.553	0.659	0.591	0.607	0.819

All correlations had at least 394 df and were statistically significant (p < 0.01).

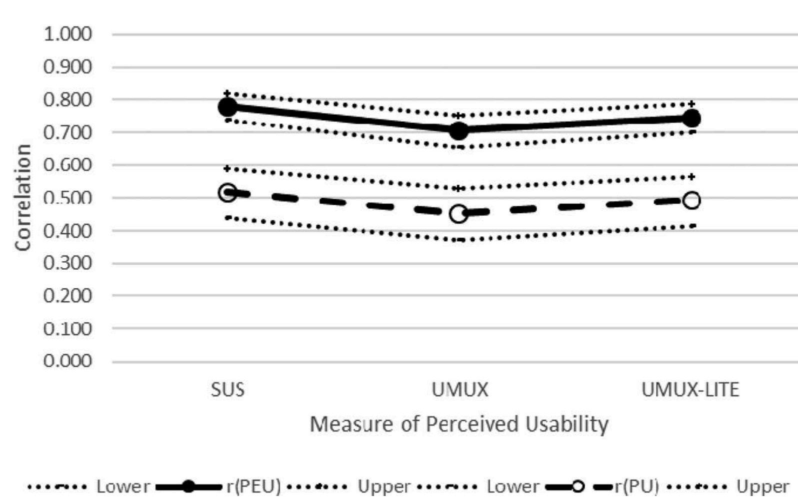

Figure 3. Correlations (with 95% confidence intervals) between perceived usability and mTAM components for Survey 2.

3.2.5. Regression analyses

Table 6 shows the adjusted coefficient of determination (R^2adj) and beta weights for regression analyses (1) using mTAM and UMUX-LITE to predict LTR and overall experience ratings and (2) substituting SUS for PEU. All values in Table 6 were statistically significant ($p < .05$).

As shown in the first row of Table 6, PU and PEU accounted for 43% of the variation in LTR (95% confidence interval ranging from 36-50%) and both beta weights were statistically significant. Replacing PU and PEU with the two items of the UMUX-LITE, as shown in the second row, produced a similar outcome, accounting for 38% of the variation in LTR (95% confidence interval ranging from 30-45%). For overall experience, the mTAM components accounted for 46% of variation in LTR (95% confidence interval ranging from 39-53%), and the UMUX-LITE items accounted for 44% (95% confidence interval ranging from 37-51%).

Although there appeared to be some reduction in the coefficient of determination when using the UMUX-LITE items in place of the mTAM components, the substantial overlap in confidence intervals indicated the differences were not statistically significant. For overall experience, as in Survey 1, the substitution of SUS for PEU produced almost identical outcomes with regard to variation accounted for and magnitude of beta weights.

Table 6. Adjusted R^2 and beta weights for regression analyses in Survey 2.

Predicting	R^2adj	Beta 1	Beta 2
LTR with PU and PEU	43%	.342	.386
LTR with UMUX01 and UMUX03	38%	.326	.382
LTR with PU and SUS	46%	.386	.394
Overall Experience with PU and PEU	46%	.271	.474
Overall Experience with UMUX01 and UMUX03	44%	.341	.420
Overall Experience with PU and SUS	49%	.330	.471

Beta weights 1 and 2 are those for the “predicting with” variables, in that order (e.g., for the second row, the beta weight for UMUX01 was 0.326 and for UMUX02 was 0.382).

3.3. Survey 3: Ratings of IBM Notes (United States)

3.3.1. Reliability

All of the measures of perceived usability had acceptable values of coefficient alpha (SUS: 0.94, UMUX: 0.91, UMUX-LITE: 0.84). The reliability of mTAM was 0.98, with 0.98 and 0.97, respectively, for PU and PEU.

3.3.2. Concurrent validity

Table 7 shows the correlations among the various measures of perceived usability and usefulness in this survey. As in Surveys 1 and 2, all correlations in Table 7 were statistically significant, but differed in magnitude. Correlations in Table 7 tended to be of higher magnitude than corresponding correlations in the other two surveys. Figure 4 shows the correlation of SUS with PEU was, as expected, significantly higher than the correlation of SUS and PU ($p < .05$), but this was not the case for the UMUX or UMUX-LITE, which had the expected pattern, but whose correlations with the components of mTAM were not significantly different.

Like Survey 1, the correlation between PU and UMUX01 (0.813) was significantly greater than PU with UMUX03

Table 7. Correlations among various measures of perceived usability and usefulness for Survey 3.

	SUS								
UMUX	0.917	**UMUX**							
UMUX-LITE	0.888	0.949	**UMUX-LITE**						
UMUX01	0.766	0.865	0.928	**UMUX01**					
UMUX03	0.881	0.895	0.927	0.720	**UMUX03**				
mTAM	0.900	0.899	0.893	0.818	0.838	**mTAM**			
PU	0.832	0.855	0.849	0.813	0.761	0.961	**PU**		
PEU	0.896	0.871	0.866	0.757	0.849	0.959	0.842	**PEU**	
OverExp	0.893	0.915	0.903	0.834	0.841	0.936	0.905	0.891	**OverExp**
LTR	0.879	0.885	0.870	0.777	0.837	0.904	0.870	0.866	0.933

All correlations had at least 539 df and were statistically significant (p < 0.01).

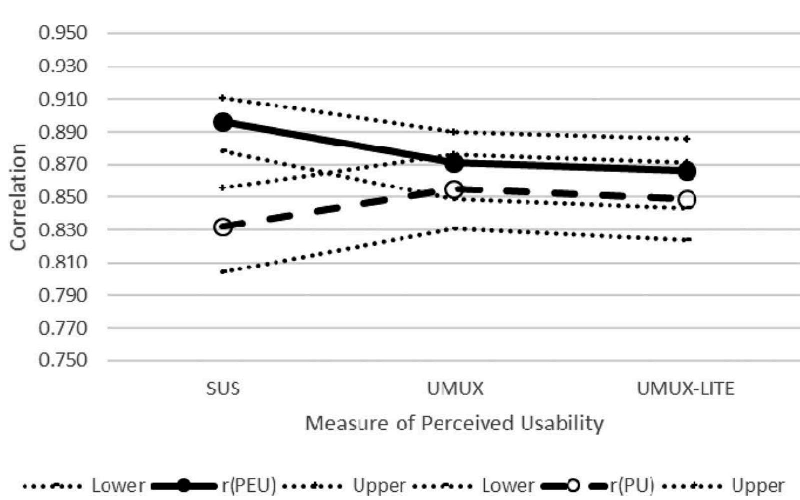

Figure 4. Correlations (with 95% confidence intervals) between perceived usability and mTAM components for Survey 3.

(0.761; $Z = 2.26$, $p = .024$). The reverse was the case for PEU, for which the correlation with UMUX01 (0.757) was significantly less than its correlation with UMUX03 (0.849; $Z = 4.34$, $p < .0001$).

3.3.3. Construct validity

The parallel analyses had the same outcomes as Surveys 1 and 2 (one-factor solutions indicated for SUS and UMUX; a two-factor solution for mTAM). For the factor analysis of the mTAM (unweighted least squares with Varimax rotation), the alignment of items with mTAM factors was as expected, with one factor composed of Items 1–6 (factor loadings ranging from 0.742 to 0.880) and one composed of Items 7–12 (factor loadings ranging from 0.706 to 0.831).

3.3.4. Regression analyses

Table 8 shows the adjusted coefficient of determination (R^2adj) and beta weights for regression analyses (1) using mTAM and UMUX-LITE to predict LTR and overall experience ratings and (2) substituting SUS for PEU. All values in Table 8 were statistically significant ($p < .05$).

PU and PEU accounted for 82% of the variation in LTR (95% confidence interval ranging from 79-84%), with both beta weights statistically significant. Replacing PU and PEU with the two items of the UMUX-LITE, as shown in the second row, produced a similar outcome, accounting for 76% of the variation in LTR (95% confidence interval ranging from 73-79%). For overall experience, the mTAM components accounted for 88% of variation in LTR (95% confidence interval ranging from 86-90%), and the UMUX-LITE items accounted for 82% (95% confidence interval ranging from 79-84%).

Although small compared to the amount of variation accounted for, using the UMUX-LITE items in place of the mTAM components resulted in statistically significant reductions in R^2adj when modeling both LTR and overall experience (nonoverlapping confidence intervals, $p < .05$). Substitution of SUS for PEU produced almost identical outcomes for the coefficient of determination and beta weights.

Table 8. Adjusted R^2 and beta weights for regression analyses in Survey 3.

Predicting	R^2adj	Beta 1	Beta 2
LTR with PU and PEU	82%	0.483	0.458
LTR with UMUX01 and UMUX03	76%	0.361	0.575
LTR with PU and SUS	83%	0.450	0.503
Overall Experience with PU and PEU	88%	0.533	0.442
Overall Experience with UMUX01 and UMUX03	82%	0.475	0.499
Overall Experience with PU and SUS	88%	0.528	0.453

Beta weights 1 and 2 are those for the “predicting with” variables, in that order (e.g., for the second row, the beta weight for UMUX01 was 0.483 and for UMUX02 was 0.458).

3.4. Correspondence with the SUS

The results of these three surveys add to the data collected so far on the correspondence of various measures of perceived usability with the SUS (most recently, Lewis, 2019b). Table 9 provides an update of this ongoing analysis of correspondence.

There are only three estimates in the table for PEU, PU and mTAM. The estimates for PEU were reasonably consistent (mean difference of 1.9 with 95% confidence interval ranging from 0.7 to 3.2), but the Slovenian estimates for PU, and consequently for mTAM, were markedly greater than the English estimates (PU: 95% confidence interval ranging from −17.3 to 27.9; mTAM: 95% confidence interval ranging from −7.1 to 14.4). For these reasons, it is too early to conduct statistical analysis to evaluate specific difference claims for the mTAM or its components.

The UMUX-LITE had the smallest absolute difference with concurrently collected SUS scores (−0.4, 95% confidence interval ranging from −2.4 to 1.6), followed by the UMUX-LITEr (0.8, 95% confidence interval ranging from −1.0 to 2.6), then the UMUX (−2.0, 95% confidence interval ranging from −4.6 to 0.7). After transformation of the mean differences in Table 9 to grade point differences using the conversion scheme in Table 1, the overall estimated differences were similar for all three UMUX-related metrics (UMUX: mean difference of −0.2 with 95% confidence interval ranging from −0.5 to 0.1, UMUX-LITE: mean difference of −0.1 with 95% confidence interval ranging from −0.4 to 0.2,

Table 9. Review of studies that have published concurrently collected measures of perceived usability and the SUS.

Study	n	SUS Version	UMUX	UMUX-LITE	UMUX-LITEr	PEU	PU	mTAM
Berkman and Karahoca (2016)*	151	Std EN	1.3	1.0	5.5	-	-	-
Borsci et al. (2015)	186	Std IT	−13.8	-	−2.9	-	-	-
Borsci et al. (2015)	93	Std IT	−12.5	-	−1.3	-	-	-
Finstad (2010)	273	Std EN	1.1	-	-	-	-	-
Finstad (2010)*	285	Std EN	0.5	-	-	-	-	-
Lewis (2018b)	618	Std EN	−0.4	−2.0	−0.2	-	-	-
Lewis (2019b: Excel)	390	Std EN	−2.3	−4.4	−1.4	-	-	-
Lewis (2019b: Word)	453	Std EN	−0.9	−2.5	1.9	-	-	-
Lewis (2019b: Amazon)*	338	Std EN	−1.6	−1.8	5.6	-	-	-
Lewis (2019b: Gmail)*	256	Std EN	0.3	0.3	4.6	-	-	-
Lewis et al. (2013)	389	Std EN	0.4	3.2	−0.7	-	-	-
Lewis et al. (2013)	402	Pos EN	2.7	3.7	0.1	-	-	-
Lewis et al. (2015)	397	Pos EN	-	5.7	1.2	-	-	-
Current study: PowerPoint	483	Std EN	−0.2	−3.5	−0.4	2.5	0.7	1.6
Current study: Gmail*	397	Std SI	−2.9	−1.9	3.6	1.5	15.8	8.6
Current study: IBM Notes	546	Std EN	−1.1	−2.5	−4.7	1.8	−0.5	0.7
Mean	**354**	**All**	**−2.0**	**−0.4**	**0.8**	**1.9**	**5.3**	**3.6**

* indicates a study in which the SUS score was a B+ or better on the Sauro–Lewis curved grading scale for the SUS (Table 1). For versions, EN = English, IT = Italian, and SI = Slovenian; Std = standard version with mixed tone, Pos = positive tone version (Sauro & Lewis, 2011). Differences shown were those obtained when subtracting other measures from the SUS.

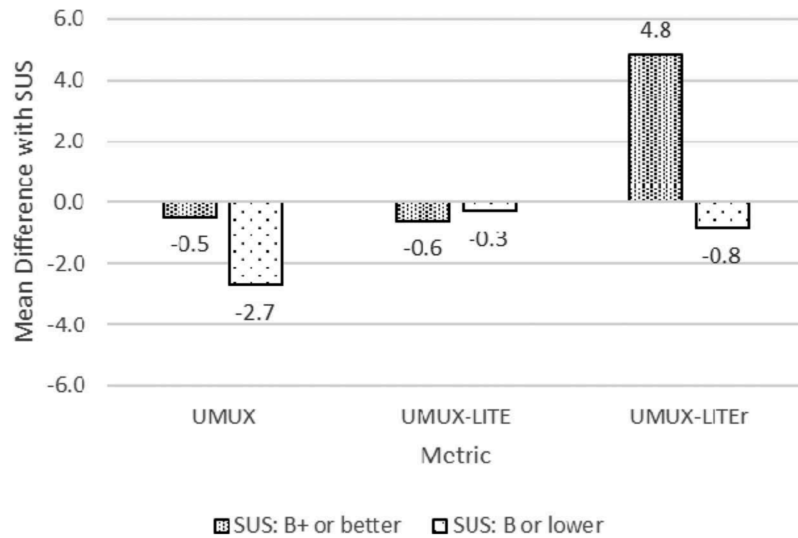

Figure 5. Mean score differences as a function of level of perceived usability.

UMUX-LITEr: mean difference of 0.1 with 95% confidence interval ranging from −0.1 to 0.4).

Figure 5 illustrates the effect of fairly high levels of perceived usability (a SUS of B+ or better on the grading scale shown in Table 1) on the magnitude of deviations between the SUS and the three UMUX-related metrics. Consistent with concerns about the possible attenuating effect of the regression formula used to compute the UMUX-LITEr, when the concurrently collected SUS mean was a B+ or better, the difference between the UMUX-LITEr mean was significantly greater than when the SUS mean was lower ($t(12) = 5.57$, $p < .0001$). That difference was not statistically significant for the UMUX ($t(13) = 0.84$, $p = .42$) or the UMUX-LITE ($t(10) = 0.16$, $p = .88$).

As shown in Figure 6, this pattern was also evident after conversion to grade points. The difference as a function of level of perceived usability was statistically significant for UMUX-LITEr ($t(12) = 4.52$, $p = .001$), but not for UMUX ($t(13) = 1.12$, $p = .28$) or UMUX-LITE ($t(10) = -0.47$, $p = .65$).

4. Discussion

In response to recent criticism of the usefulness of the construct of usability, we investigated the relationships between measures of perceived usability and the components of a modified version of the Technology Acceptance Model (mTAM) – Perceived Usefulness (PU) and Perceived Ease-of-Use (PEU). In three surveys, respondents used SUS, UMUX-LITE and mTAM to rate their actual (as opposed to expected) experience with a popular commercial software application, an online e-mail system, and an internal e-mail system, with variation in the background and languages of the populations surveyed.

4.1. Psychometrics

All metrics used in the surveys – SUS, UMUX, UMUX-LITE and mTAM (including its PU and PEU subscales) had acceptably high levels of reliability. The correlations observed

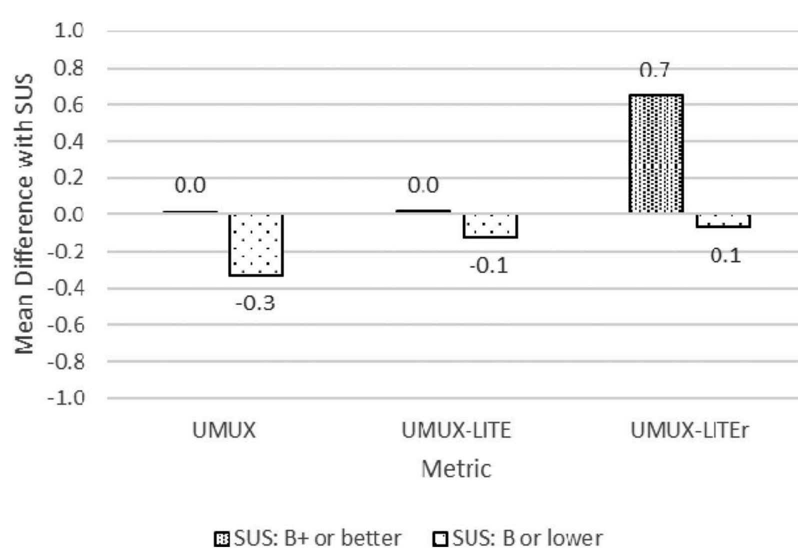

Figure 6. Mean grade point differences as a function of level of perceived usability.

across all three surveys (see Tables 3, 5 and 7) indicated acceptably high and statistically significant levels of concurrent validity (all $r > .40$, $p < .01$).

Parallel analyses in all three surveys indicated retention of one factor for the SUS and UMUX, and retention of two factors for the mTAM. The factor structures of the mTAM in Surveys 1 and 3 were exactly as expected, with the alignment of Items 1–6 on one factor and 7–12 on the other. Its factor structure in Survey 2 was close, with the exception of mTAM06 which aligned with Items 7–12 instead of Items 1–5, suggesting a possible issue with its translation.

Figures 2–4 show the results of convergent and divergent validity analyses for the three surveys. The expectation was that the three measures of perceived usability, SUS, UMUX and UMUX-LITE, would correlate significantly more with the mTAM component of PEU (perceived ease-of-use) than they would with PU (the mTAM component of perceived usefulness). This expectation was met for the SUS in all three surveys, and was met for the UMUX and UMUX-LITE in Surveys 1 and 2, but not in Survey 3. The UMUX and UMUX-LITE results in Survey 3 followed the expected pattern, but were not statistically significant.

Because the items of the UMUX-LITE bear a strong resemblance to the PU and PEU components of the mTAM, the expectation was that the first item of the UMUX-LITE (UMUX01: capabilities meet requirements) would more strongly correlate with PU and the second UMUX-LITE item (UMUX03: easy to use) would correlate more strongly with PEU. The results were as expected for Surveys 1 and 3, and as expected for PEU in Survey 2. In Survey 2, the correlation between PU and UMUX01 was not significantly greater than the correlation of PU with UMUX03.

The majority of the deviations from expectation were in Survey 2, for which all questionnaires with the exception of the SUS were newly translated. Despite the care taken in the translation process, it is possible that there might have been an issue with the translation of mTAM06 that affected its alignment in the factor analysis and correlation of PU with UMUX01. For the discrepant results in Survey 3, it is important to note that the differences were in the expected direction but were not statistically significant, possibly due to ceiling effects brought on by the very high magnitudes of the correlations in question.

In summary, these results indicate that the measures used in these surveys had acceptable levels of the basic psychometric properties of reliability and concurrent validity. With a few minor exceptions (alignment of mTAM06 in factor analysis and correlations between mTAM components and UMUX01 in Survey 2, correlations of UMUX and UMUX-LITE with mTAM components in Survey 3) the results for construct validity, convergent and divergent validity were as expected.

4.2. Regression

4.2.1. Regression with components of mTAM

All regression models predicting likelihood-to-recommend and overall experience with PU and PEU were statistically significant, with reasonably consistent results for both models from each survey. The magnitudes of the coefficients of determination appeared to differ as a function of study (no overlap of confidence intervals, so $p < .05$ for each pair of studies), with the highest levels for Study 3 (IBM employees rating IBM Notes), then Study 1 (IBM employees rating PowerPoint), then Study 2 (Slovenian educators, students and employees rating Gmail). A potential explanation of these differences is that in Study 3, participants rated an application that they had no choice in using (IBM Notes), leaving little room for factors other than PU and PEU to drive variation in the outcome metrics. In Study 2, participants were not required to use Gmail, so factors other than PU and PEU may have been in play when they made their choice to use the application. In Study 1 participants were not required to use PowerPoint, but their options for presentation software were limited, so it is reasonable that the amount of variation accounted for in Survey 1's regression models was between that of Surveys 2 and 3.

4.2.2. Substituting SUS for PEU

The regression models were similar with regard to the magnitudes of coefficients of determination and beta weights in all three surveys when substituting the SUS for PEU. As Lewis (2018b) 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. These regression results provide evidence that the PEU component of the mTAM appears to be another measure of the construct of perceived usability, connecting the TAM to the construct of perceived usability through the mTAM and providing evidence against the claim that the construct of usability is a theoretical dead end (Tractinsky, 2018).

4.2.3. Substituting UMUX-LITE items for mTAM components

Substituting UMUX01 and UMUX03 for PU and PEU demonstrated that, with a few exceptions, the regression models were similar. For all surveys, all four models had statistically significant coefficients of determination and beta weights. For all surveys, the coefficients of determination were slightly smaller when modeling with UMUX-LITE items, but this drop in variation accounted for was only significant in Survey 3. It is possible that the significance obtained in Survey 3 was due to it having the largest sample size and the largest (and therefore least variable) coefficients of determination. The consistency of outcomes for these regression models provides additional support for the use of the UMUX-LITE as a concise UX metric with theoretical and empirical connections to the TAM.

4.3. Correspondence of UMUX-LITE with the SUS

SUS and UMUX-LITE measures were reasonably consistent when considering mean raw differences, and were very consistent for mean grade point differences. These findings support the use of the UMUX-LITE as a concise UX metric that can be interpreted using the Sauro–Lewis curved grading scale (Table 1). The data also indicate that researchers and practitioners should prefer the unadjusted UMUX-LITE over the regression-adjusted UMUX-LITEr because the UMUX-LITE closely corresponds to the SUS not only for average levels of perceived usability but also for higher levels where the UMUX-LITEr tends to fall short of matching the SUS (Figures 5 and 6).

4.4. Limitations and future research

Much of the data in the current literature that examines the relationship among various measures of perceived usability (e.g., Lewis, 2018b, 2019b; Lewis et al., 2013, 2015), as well as the current research, are from surveys rather than usability studies. Generalizability would be enhanced with the addition of more data from traditional usability studies, keeping in mind the difficulty of acquiring sufficiently large sample sizes from this method.

Despite the effort in the current research to collect data from multiple sources and for multiple products, replication of this work with other user populations and products, performed by other researchers, could enhance the generalizability of the findings. Of particular interest would be variation in the extent to which users have or do not have a choice in the use of the product and its effect on the magnitude of variance accounted for in regression models.

The body of research on the correspondence between the SUS and UMUX-LITE has grown over the past few years, and across the 16 assessments to date (Table 9) the confidence intervals around the mean differences appear to be reasonably narrow, especially after conversion to grade points. On the other hand, until researchers have replicated these outcomes across an even wider variety of systems and research methods, practitioners should be appropriately cautious when using the Sauro–Lewis curved grading scale to interpret the UMUX-LITE.

The Slovene translations of the UMUX and mTAM were reasonably successful in the current research, but it might be possible to improve the translations of UMUX01 and mTAM06 in future research. There would also be value in replicating this work in other languages.

5. Conclusions

Expanding upon the use of the original TAM to predict likelihood of future use, the results of these surveys have shown a modified version of the TAM (mTAM) is predictive of likelihood-to-recommend and overall experience, a finding of interest to UX researchers. Substitution of the SUS for PEU in the models did not affect the magnitudes of coefficients of determination or beta weights, supporting the hypothesis that PEU and SUS are independently developed metrics that measure the same construct of perceived usability. As predicted, the SUS always correlated more highly with PEU than PU, and it was usually the case that the items of the UMUX-LITE correlated as expected with the parameters of the mTAM.

Of critical importance to UX practitioners, UMUX-LITE scores appear to have high correspondence with concurrently collected SUS scores for both standard scores and after conversion to grade point averages using the Sauro–Lewis curved grading scale. For these reasons, UX researchers and practitioners should strongly consider using the UMUX-LITE as a concise measure of perceived usability. For practitioners who plan to switch from SUS to UMUX-LITE, we recommend a period of concurrent data collection to ensure an acceptable level of correspondence.

The main theoretical outcome of this research was evidence from regression models that the PEU component of the mTAM appears to be another measure of the construct of perceived usability, connecting the TAM to the construct of perceived usability through the mTAM and providing evidence against the claim that the construct of usability is a theoretical dead end.

ORCID

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

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. doi:10.1080/10447310802205776

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

Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. *Journal of Usability Studies*, *11*(3), 89–109.

Blažica, B., & Lewis, J. R. (2015). A slovene translation of the system usability scale: The SUS-SI. *International Journal of Human-Computer Interaction*, *31*(2), 112–117. doi:10.1080/10447318.2014.986634

Borsci, S., Federici, S., Gnaldi, M., Bacci, S., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: An exploratory analysis of SUS, UMUX and UMUX-LITE. *International Journal of Human-Computer Interaction*, *31*, 484–495.

Borsci, S., Federici, S., Malizia, A., & De Filippis, M. L. (2019). Shaking the usability tree: Why usability is not a dead end, and a constructive way forward. *Behaviour & Information Technology*, *38*(5), 519–532. doi:10.1080/0144929X.2018.1541255

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). London, UK: Taylor & Francis.

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

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human–computer interface. In E. Soloway, D. Frye, & S. B. Sheppard (Eds.), *Proceedings of CHI 1988* (pp. 213–218). Washington, DC: Association for Computing Machinery.

Davis, D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*, 319–339. doi: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. doi: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). New York, NY: Association for Computing Machinery.

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

Finstad, K. (2013). Response to commentaries on "the usability metric for user experience". *Interacting with Computers*, *25*, 327–330. doi:10.1093/iwc/iwt005

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

Kirakowski, J., & Dillon, A. (1988). *The computer user satisfaction inventory (CUSI): Manual and scoring key*. Cork, Ireland: Human Factors Research Group, University College of Cork.

Lewis, J. R. (1990). *Psychometric evaluation of a post-study system usability questionnaire: The PSSUQ* (Tech. Report 54.535). Boca Raton, FL: International Business Machines Corp. doi:10.1099/00221287-136-2-327

Lewis, J. R. (2018a). Is the report of the death of the construct of usability an exaggeration? *Journal of Usability Studies*, *14*(1), 1–7.

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

Lewis, J. R. (2018c). The system usability scale: Past, present, and future. *International Journal of Human-Computer Interaction*, *34*(7), 577–590. doi: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. doi:10.1080/10447318.2018.1533152

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). Paris, France: 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. doi:10.1080/10447318.2015.1064654

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Pearson, S. W., & Bailey, J. E. (1980). Measurement of computer user satisfaction. *ACM SIGMETRICS Performance Evaluation Review*, *9*(1), 59–68. doi:10.1145/1041872

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

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: Association for Computing Machinery.

Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of CHI 2011* (pp. 2215–2223). Vancouver, Canada: ACM.

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

Tractinsky, N. (2018). The usability construct: A dead end? *Human-Computer Interaction*, *33*(2), 131–177. doi:10.1080/07370024.2017.1298038

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

About the Authors

Urška Lah is a PhD student of informatics at the University of Maribor, Faculty of Electrical Engineering and Computer Science (Diploma in informatics, 2009), and an information systems developer in EMO-Orodjarna's information and communications technology department. Her research interest is human-computer interaction.

James R. Lewis is a distinguished user experience researcher at MeasuringU, an IBM Master Inventor emeritus (over 90 patents), and the author of Practical Speech *User Interface Design* and (with Jeff Sauro), *Quantifying the User Experience*.

Boštjan Šumak is an assistant professor and researcher at the University of Maribor, Faculty of Electrical Engineering and Computer Science (PhD, 2011), with over 18 years of experience in research and teaching at the institution. His research interests include human-computer interaction, user experience, brain-computer interfaces, and information technology acceptance.

Appendix

This appendix documents the English versions of the standardized questionnaires used in these surveys.

The System Usability Scale (SUS) is a questionnaire commonly used to assess perceived usability.

Again, as you answer these questions please think about all the tasks that you do with this product.

Please read each statement and indicate how strongly you agree or disagree with the statement. Please read the statements carefully, but don't spend a lot of time on an item -- your first impression is fine. If you do not have an opinion about a statement, please select 3 (the middle point) rather than leaving it blank.

Note that for this questionnaire (SUS), the tone of the statements is mixed -- half are positive statements and half are negative (alternating positive and negative from start to end), so be sure to take that into account when you provide your rating.

Also note that in this questionnaire a 1 indicates strong disagreement and a 5 strong agreement -- the mixed tone of the items affects whether a low or high rating indicates a poor or good user experience.

Please indicate the extent to which you agree with the following statements **where 1 = Strongly disagree and 5 = Strongly agree.**

	1	2	3	4	5
1. I think that I would like to use this product frequently.	○	○	○	○	○
2. I found the product unnecessarily complex.	○	○	○	○	○
.3 I thought the product was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this product.	○	○	○	○	○
5. I found the various functions in the product were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this product.	○	○	○	○	○
7. I would imagine that most people would learn to use this product very quickly.	○	○	○	○	○
8. I found the product very awkward to use.	○	○	○	○	○
9. I felt very confident using the product.	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this product.	○	○	○	○	○

Figure A1. The System Usability Scale (SUS).

The Usability Metric for User Experience (UMUX) is another questionnaire designed to assess perceived usability.

Again, please think about all the tasks that you do with this product while you answer these questions.

Please read each statement and indicate how strongly you agree or disagree with the statement. Please read the statements carefully, but don't spend a lot of time on an item -- your first impression is fine. If you do not have an opinion about a statement, please select 4 (the middle point) rather than leaving it blank.

Note that for this questionnaire (UMUX), the tone of the statements is mixed -- half are positive statements and half are negative (alternating positive and negative from start to end), so be sure to take that into account when you provide your rating.

Also note that in this questionnaire a 1 indicates strong disagreement and a 7 strong agreement -- the mixed tone of the items affects whether a low or high rating indicates a poor or good user experience.

Please indicate the extent to which you agree with the following statements **where 1 = Strongly disagree and 7 = Strongly agree.**

	1	2	3	4	5	6	7
1. This product's capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using this product is a frustrating experience.	○	○	○	○	○	○	○
3. This product is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with this product.	○	○	○	○	○	○	○

Figure A2. Usability Metric for User Experience (UMUX).

The Technology Acceptance Model (TAM) is designed to give you an opportunity to rate this product's usefulness and ease-of-use.

To as great an extent as possible, think about all the tasks that you do with the product while you answer these questions.

Please read each statement and indicate how strongly you agree or disagree with the statement. Please read the statements carefully, but don't spend a lot of time on each item -- your first impression is fine.

Note that for this questionnaire (TAM), all items have a positive tone so greater levels of agreement (to the right of the scale) indicate a better user experience.

Please indicate the extent to which you agree with the following statements **where 1 = Extremely disagree and 7 = Extremely agree.**

	1	2	3	4	5	6	7
1. Using this product in my job enables me to accomplish tasks more quickly than other products in its class.	○	○	○	○	○	○	○
2. Using this product improves my job performance.	○	○	○	○	○	○	○
3. Using this product in my job increases my productivity.	○	○	○	○	○	○	○
4. Using this product enhances my effectiveness on the job.	○	○	○	○	○	○	○
5. Using this product makes it easier to do my job.	○	○	○	○	○	○	○
6. I have found this product useful in my job.	○	○	○	○	○	○	○
7. Learning to operate this product was easy for me.	○	○	○	○	○	○	○
8. I found it easy to get this product to do what I want it to do.	○	○	○	○	○	○	○
9. My interaction with this product has been clear and understandable.	○	○	○	○	○	○	○
10. I found this product to be flexible to interact with.	○	○	○	○	○	○	○
11. It was easy for me to become skillful at using this product.	○	○	○	○	○	○	○
12. I found this product easy to use.	○	○	○	○	○	○	○

Figure A3. Modified Technology Acceptance Model (mTAM).