Computers in Human Behavior Reports 22 (2026) 101014

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Full length article

Inventory of User Expectations for Technology (iExpect)

Steeven Villa [a,*], Thomas Kosch [b], Agnes Mercedes Kloft [c], Jasper Quinn [c], Sari Kujala [c], Robin Welsch [c]

[a] LMU Munich, Frauenlobstrasse 7a, Munich, 80337, Germany
[b] HU Berlin, Rudower Chaussee 25, Berlin, 12489, Germany
[c] Aalto University, Konemiehentie 2, Espoo, 02150, Finland

ARTICLE INFO

Keywords:
User expectations
Scale development
Technology evaluation
Human-Computer Interaction (HCI)
Expectation bias

ABSTRACT

User expectations of novel interactive technologies are critical during their evaluation. Prior research shows that matching expectations improves experience and interaction, while a mismatch can lead to disappointment. We developed the Inventory of User Expectations for Technology (iExpect), which measures user expectations before interaction. To develop the scale, we first created a list of items grounded in the Technology Acceptance Model and the existing literature which were reviewed by 11 experts. Next, we performed a study with 259 participants and used exploratory factor analysis to reduce the scale to 22 items (three factors: Anticipated Ease of Use, Anticipated Usefulness, Anticipated Enjoyment). Subsequently, we confirmed its test–retest reliability and construct validity with a study of 278 participants. The iExpect scale enables researchers and practitioners to measure user expectations before interaction and assists designers, user experience professionals and researchers in considering technology expectations and mitigating for expectation-based effects in technology evaluation.

1. Introduction

Consider a user encountering a new technology advertised as “revolutionizing productivity”, such claims shape expectations immediately, influencing perceptions before any actual interaction occurs. Emerging technologies such as Artificial Intelligence (AI) and Human Augmentation (HA) amplify this process by generating strong anticipatory beliefs about functionality and usability.

Expectations shaped by media and word-of-mouth influence users’ mental models, affecting how they evaluate and interact with the technology (Pataranutaporn et al., 2023). In many cases, these expectations can override actual experiences, leading to phenomena such as placebo effects (Kosch et al., 2022), increased risk-taking (Villa, Kosch et al., 2023), or disproportionately positive responses in qualitative evaluations, regardless of the technology’s objective quality (Brown et al., 2014).

Thus, the core objective of this research is to develop and validate a psychometrically sound instrument that systematically measures users’ anticipatory beliefs toward interactive technologies before first use, in order to better understand and control expectation-driven biases in technology evaluation. Given the influence of user expectations on interactions with new technologies and products (Brown et al., 2014), it is crucial to measure these expectations as a control variable in technology evaluation.

Traditional methods focus on expectation-experience and disconfirmation models. These approaches often rely on custom questions that are subsequently modeled using polynomial methods or questions derived from the Technology Acceptance Model (TAM) (Brown et al., 2008). Similarly, in HCI research, questions are typically derived from the TAM and its associated dimensions, supplemented by instruments such as the System Usability Scale (SUS) (Lewis, 2018) or models such as the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003). The primary limitation of these approaches is that they may not adequately capture the specific nature of user expectations prior to actual use, as these instruments are primarily designed to measure post-interaction experiences rather than pre-interaction expectations (Kujala et al., 2017). This temporal gap represents a critical theoretical boundary: while TAM and UTAUT are well suited for predicting adoption based on experience-driven perceptions, no validated instrument systematically measures the anticipatory beliefs that precede and potentially bias these perceptions. Most importantly, these measures were not psychometrically validated.

The theoretical basis of this work builds on expectation-disconfirmation theory, technology acceptance models, and research on placebo and expectancy effects in human-technology interaction. Unlike TAM or UTAUT, which focus on post-use perceptions, our

* Corresponding author.
E-mail address: steeven.villa@ifi.lmu.de (S. Villa).

https://doi.org/10.1016/j.chbr.2026.101014
Received 12 May 2025; Received in revised form 5 February 2026; Accepted 13 March 2026
Available online 18 March 2026

approach targets pre-use cognitive and affective beliefs that shape subsequent evaluations.

To address this gap, we present the Inventory of User Expectations for Technology (iExpect) scale, a tool specifically designed to measure user expectations toward interactive technologies before their use. Our approach contributes to a more transparent and robust evaluation framework for new technologies (Denisova & Cairns, 2019; Kloft et al., 2024; Kosch et al., 2022; Kujala & Miron-Shatz, 2015; Kujala et al., 2017; Villa, Kosch et al., 2023). In developing this scale, we followed a structured approach based on the guidelines outlined by Boateng et al. (2018) for scale development and evaluation.

We began by collecting items from related instruments, including derivations of TAM and UTAUT, resulting in an initial pool of 329 items. Through a multi-step refinement process conducted by the research team, this pool was reduced to 89 items. We then involved 11 experts to assess the item pool, leading to a further reduction to 58 items. To refine the scale further, we conducted a study with 259 participants and conducted a confirmatory factor analysis to reduce the number of items. Followed by a confirmatory factor analysis (CFA) with an additional sample of 278 participants we confirmed the scale structure, resulting in a three-factor (Anticipated Ease of Use; EU, Anticipated Usefulness; US, and Anticipated Enjoyment; AE) scale comprising 22 items. Finally, we assessed the scale's temporal stability and evaluated its discriminant, convergent, and concurrent validity, ensuring its robustness and generalizability.

This instrument is strategically situated in the pre-use phase, yet it is intended to inform and improve evaluations during the initial stage of product use (for after-use phases, researchers can still use already available instruments), thereby offering valuable insights for the early phase of technology development and deployment, controlling for potential expectation-based effects (e.g., placebo effects or expectation-dis/confirmation). In this context, we conceptualize Anticipated Enjoyment (AE) as a distinct from previous frameworks pre-use construct, capturing expected pleasure before interaction, positioning it as a potential source of evaluative bias rather than an experiential outcome.

2. Related work

Previous work investigated how user expectations shape perceptions of technologies. This section reviews relevant literature regarding user expectations and the associated models.

2.1. Models of expectations

Users' experiences with new technologies are influenced by their initial assumptions; Raita and Oulasvirta (2011) define product expectations as "*beliefs and emotions related to a product that is formed before its actual use*" underscoring that factors such as advertisements, branding, word of mouth, product reviews, discussion forums, and exposure to similar products play a crucial role in shaping user expectations. These expectations, in turn, significantly impact user experiences (den Ouden et al., 2006; Kujala et al., 2017). When expectations are not met, use is accompanied by user dissatisfaction (Bly et al., 2006; den Ouden et al., 2006; Olsson & Salo, 2012). Consequently, it is essential to understand user expectations of technology prior to interaction.

Several theoretical models have been developed in psychological research to understand and model expectations. Among these, the most relevant, as identified by Brown et al. (2014), include the *assimilation model*, the *contrast model*, the *generalized negativity model*, the *assimilation-contrast model*, and the *Expectations Only model*: *The assimilation model* is grounded in cognitive dissonance theory, as proposed by Festinger (1962). Cognitive dissonance theory suggests that a deviation from one's expectations creates a state of discomfort known as dissonance. To alleviate this discomfort, individuals tend to adjust their subsequent evaluations of outcomes to align more closely with their original expectations. On the other hand, the *Contrast Model* is based on the disconfirmation of expectations theory. This theory posits that the final evaluation of an outcome is influenced by the direction and magnitude of the gap between what was expected and what was experienced. Unlike cognitive dissonance theory, which focuses on reconciling expectations, disconfirmation of expectations theory emphasizes the impact of the difference between expectations and subsequent evaluations, often resulting in an outcome evaluation biased toward the experience.

The *generalized negativity model* is informed by the met expectations hypothesis, as discussed by Irving and Meyer (1994) and Porter and Steers (1973). According to this hypothesis, any discrepancy between expectations and experiences — whether the expectations are exceeded or unmet — leads to a decrease in outcome evaluation (Olson & Dover, 1979). The *assimilation-contrast model* integrates principles from cognitive dissonance theory and expectation disconfirmation theory (Anderson, 1973). This model posits that when the difference between expectations and experiences is minimal, outcome evaluations align with the original expectations. However, when the difference is substantial, the evaluation becomes more influenced by the experience due to the stronger contrast (Klein, 1999).

Finally, the expectations-only model suggests that expectations alone directly predict outcomes, with no influence from actual experiences (Brown et al., 2008). This model represents a perfect assimilation toward an individual's initial beliefs, although a complete alignment between expectations and experiences may never be fully realized. For instance, Pulakos and Schmitt (1983) found a positive relationship between expectations and satisfaction, whereas Miceli (1987) argued that the expectations-only model is more likely to apply when expectations are assessed early. Over time, as experiences accumulate, the influence of initial expectations may diminish, leading to a shift toward an experiences-only model.

Research in technology evaluation has generally supported the assimilation theory, suggesting that expectations shape experiences (Hartmann et al., 2008; Kujala et al., 2017; Raita & Oulasvirta, 2011; van Schaik & Ling, 2008). Moreover, two studies specifically challenged the contrast theory by showing that negative expectations resulted in lower user ratings compared to positive ones, even though negative expectations are more easily surpassed (Hartmann et al., 2008; Raita & Oulasvirta, 2011). On the other hand, Michalco et al. (2015) provided evidence for both assimilation and contrast theories. In their study, participants played a game, and both primed and naturally occurring expectations influenced user experience ratings afterward.

Kujala et al. (2017) conducted a longitudinal study demonstrating that expectations significantly influence early usability and user experience evaluations. They caution against relying solely on initial evaluations of new products, suggesting that continuous user feedback collection is essential. In addition, Kujala and Miron-Shatz (2015) recommended using expectations as a benchmark to better interpret early evaluation results. In agile processes and prototype development of interactive technology, conducting a longitudinal, multi-stage evaluation with the same users is often impractical. To address this limitation, controlling for user expectations can enhance the accuracy of technology assessment in these contexts.

2.2. Heightened user expectations in interactive systems

In interactive systems, heightened user expectations toward a not-yet-used system or interface can alter subjective and objective perceptions of that system. Wells et al. (2010) demonstrated that expectations significantly influence the adoption of information technology innovations. Similarly, Boot et al. (2013) along with others (Kloft et al., 2024; Kosch et al., 2022; Villa, Kosch et al., 2023) recommend measuring user expectations in experimental settings before and after using a system to avoid perceived improvements through elevated user expectations. Boot et al. (2013) states that differing expectations can lead to

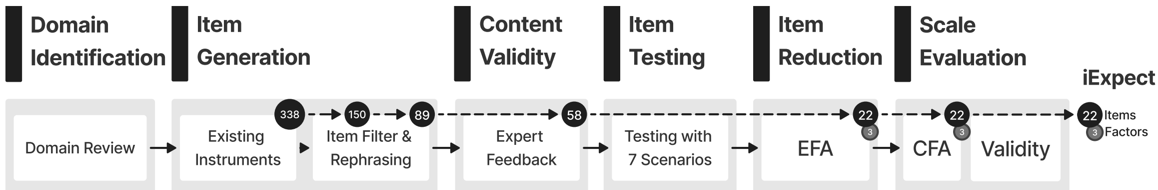

Fig. 1. Process overview: The creation of the iExpect was performed in accordance with the best practices suggested by Boateng et al. (2018). The numbers inside the black circles (●) describe the number of items in each stage.

erroneous conclusions in intervention studies, which can be mitigated by measuring expectations and employing alternative experimental designs to control for user expectations and improve the reliability of results. In essence, heightened user expectations can lead to more favorable evaluations of technology, even when it is not objectively superior.

Recent research has examined the impact of user expectations in AI systems, where users encounter similar challenges related to functional transparency. Kosch et al. (2022) found that user expectations are altered when users are primed with the description of an adaptive AI system despite interacting with a non-adaptive system. Participants believed they performed better when they thought they were using an adaptive AI system. Villa, Kosch et al. (2023) explored how placebo cognitive augmentation systems influence users' risk-taking behavior, discovering that participants took more risks during sham augmentations. Pataranutaporn et al. (2023) studied how participants perceived voice agents with different intents including neutral, malevolent, and benevolent. The results showed that participants rated the benevolent voice agent as more helpful, even though all conditions involved a neutral voice agent. Kloft et al. (2024) investigated the impact of heightened AI expectations on performance through user expectations, finding that participants performed better when they believed an AI was enhancing their task interface, regardless of the AI's actual presence. This highlights the strong influence of user expectations on AI interactions and evaluations. Further research by Bosch et al. (2024) showed that the narrative around using a specific display refresh rate can significantly skew participants' performance perceptions.

Consequently, user expectations change how users assess their performance, objective task efficiency, and overall system, thus undermining efforts in technology evaluation. Thus, the measurement of user expectations in psychology studies was proposed in 2013 (cf. Boot et al. (2013)). Understanding these expectations is vital because they influence how individuals perceive, engage with, and evaluate technology. However, despite this recognition, technology evaluation has largely overlooked this aspect, potentially due to the absence of standardized methods to capture and quantify such expectations. This work addresses this circumstance by proposing a questionnaire that assesses user expectations toward novel technologies.

3. Constructing the technology expectations questionnaire

Over the last five years, a notable trend has emerged in technology evaluation research: the development of custom measurement instruments tailored to specific research needs, as opposed to relying solely on traditional psychometric tools from fields such as psychology and economics (Bentvelzen et al., 2021; Schmidmaier et al., 2024; Villa, Niess et al., 2023; Woźniak et al., 2021). To ensure rigor in this process, we have adhered to the best practices for questionnaire development as recommended by Boateng et al. (2018) (see Fig. 1).

4. Item generation

We approached the item generation in four steps: First, we systematically identified relevant instruments through a literature search targeting scales related to technology experience and user expectations. The selection criteria required that instruments (a) measured constructs conceptually related to user expectations or technology perceptions, (b) were empirically validated, and (c) contained items applicable to pre-use or early-stage technology evaluation contexts. Second, we extracted all items from the selected instruments, resulting in an initial pool of 329 items. Third, three researchers independently filtered this pool using conservative criteria based on relevance to pre-use expectations and conceptual similarity to avoid redundancy. Fourth, five researchers collaboratively evaluated the remaining items based on (a) relevance to the target construct, (b) clarity of formulation, and rephrased items or suggested removals where necessary.

In detail, we conducted a domain identification by collecting already existing instruments that captured a construct similar to the construct of interest: user expectations. A search in databases such as the *ACM Digital Library*, *IEEE Explore*, *ArXiv*, and *Web of Science* was conducted with the following search terms, `(expec*) AND (technolog*)`, until saturation was reached (similar items were constantly found in the search).

As a result of this process, the following established questionnaires were included: the Technology Acceptance Model (TAM) (Davis, 1985), the Unified Theory of Acceptance and Use of Technology (UTAUT (Madigan et al., 2017; Maillet et al., 2015; Venkatesh et al., 2003) & UTAUT 2.0 (Venkatesh et al., 2012)), the User Experience Questionnaire (UEQ) (Schrepp et al., 2017), the System Usability Scale (SUS (Bangor et al., 2008) & the positive SUS (Lewis, 2018)), the Perceived Creepiness of Technology Scale (PCTS (Woźniak et al., 2021)), and the modular evaluation of key Components of User Experience (meCUE (Minge et al., 2017)). Additional questionnaires from the model Motivation, Engagement and Thriving in User Experience (METUX (Peters et al., 2018)), namely the Autonomy & Competence in Technology Adoption Questionnaire (ACTA) and the Technology-based Experience Need Satisfaction (TENS) with its adaptations for the interface (TENS-Interface), the task (TENS-task), and life (TENS-Life), were also incorporated.

We collected a total of $N = 329$ items from the instruments listed above; the breakdown of the item number contributed by each instrument to the overall pool is listed in Table 1. During the item generation process, the researchers prioritized the relevance of the items in measuring technology expectancy prior to interaction, not posterior evaluations.

Then, a team of three researchers (1) rephrased the pool of items and (2) ranked and filtered the items based on their potential to measure technology expectancy. All three researchers were either PhD-level researchers or professors with extensive experience (more than three years) in HCI, user experience research, and technology evaluation, ensuring familiarity with the theoretical models underlying the selected scales. Items were rephrased, preserving a high similarity to the original scales while ensuring they remained technology-agnostic, thus capable of evaluating technologies before use in the broader sense. For the ranking and filtering, the criteria included relevance, potential duplication, and similarity after translation/rephrasing. Having conservative criteria, items that were literal duplicates, highly similar or out of scope were marked for removal. Then the researchers reviewed the remaining items, rating them on a scale from 1 to 5, where 1 was high relevance and 5 was low relevance. Items ranked 3 to 5 were excluded, narrowing

Table 1
Summary of questionnaires included. Number of items by instrument with references.

Instrument	Author(s)	Items
ACTA	Peters et al. (2018)	14
UTAUT	Madigan et al. (2017)	16
Positive SUS	Lewis (2018)	10
UTAUT	Maillet et al. (2015)	40
PCTS	Woźniak et al. (2021)	8
UTAUT 2.0	Venkatesh et al. (2012)	25
TAM	Davis (1985)	31
meCUE 2.0	Minge et al. (2017)	35
TENS	Peters et al. (2018)	37
SUS	Bangor et al. (2008)	10
UTAUT	Venkatesh et al. (2003)	31
UEQ	Schrepp et al. (2017)	72
	Total	329

the list to 150 items. This selection process focused on retaining only the most pertinent items for the second-to-last cut.

To refine and finalize the initial pool of items for the iExpect questionnaire, the set of 150 items was evaluated by a team of five researchers. This team included three researchers who were involved in the previous step and two researchers who were not familiar with the initial item pool. Through this evaluation process, the list was reduced to 89 items, which constituted the initial pool for the iExpect questionnaire.

5. Content validity

Following the creation of the item pool, the next step involves assessing the content validity of the items to ensure they accurately measure the intended target domain. In line with Boateng et al. (2018)'s recommendations, we evaluated the content validity of the 89-item pool by soliciting external expert feedback. External experts reviewed each item and rated its relevance to the construct, providing detailed feedback to refine, rephrase, add, or remove items to the pool. To measure the content validity of the pool of items, we calculated the Content Validity Index (CVI) (Polit & Beck, 2006; Wynd et al., 2003).

5.1. External expert feedback

We conducted an online survey with eleven experts to calculate the Content Validity Index (CVI). The experts were asked to rate each of the 89 items on a 4-point scale based on its relevance to the topic.

Procedure Participants were informed of the study's purpose, with particular emphasis on the items being potential candidates for a scale aimed at measuring *A Priori Expectations of Novel Technology* from the user's perspective. To ensure consistent interpretation of the scale's purpose, the construct definition was displayed on every page of the survey (see Definition #1). Participants were instructed to rate the relevance of each of the 89 initial items for inclusion in the questionnaire. The items were presented in a randomized order, each accompanied by a question regarding its relevance to the intended construct. Experts rated each item on a 4-point scale: Not Relevant, Somewhat Relevant, Quite Relevant, or Very Relevant. This rating procedure was based on previous work in scale development (Schmidmaier et al., 2024) and content validation research (Polit & Beck, 2006; Wynd et al., 2003).

Additionally, experts were given the option to provide comments on each item in a free text field. Finally, participants were asked to self-assess their expertise in related areas using 5-point scales and to provide demographic information.

Definition #1 (A Priori Expectations of Novel Technology): These are the anticipated functionalities or perceived usefulness of a technology or system that a user has not yet experienced. Since the user lacks firsthand experience or objective data about the technology, these expectations are not based on informed assessments. Instead, these assessments are constructed on assumptions, public opinions, media, marketing, or hype surrounding the technology or system. In this sense, it differs from the Expectation-Experience models as this concept is intended to capture the prior expectations of a technology and not the mismatch between a priori and a posteriori assessment of a technology.

Participants We conducted a study involving 11 experts in Human-Computer Interaction (HCI), Psychology, and Questionnaire Development. These experts were selected based on their publication records within these domains and their contributions to the ACM CHI Conference. Out of the 20 experts who were invited to participate, 11 accepted the invitation. Participation in this study was entirely voluntary, with no compensation provided. The average time required for each expert to review the complete item list was approximately 33 min ($M = 33.17, SD = 12.43$). Of the 11 experts, 5 identified as female, and 6 as male. Participants had an average age of 33 years old ($M = 33.09, SD = 3.01$). A detailed summary of the experts' backgrounds is provided in Table 2.

Data analysis We calculated the Content Validity Index (CVI) for each item based on the ratings provided by the experts. According to the guidelines proposed by Polit and Beck (2006), a threshold of 0.57 is deemed acceptable. For items that fell within a borderline range of 0.55 to 0.65, we conducted further inspections to ensure their validity. Following a consensus among the research team, considering both expert feedback and the calculated CVI values, a total of 58 items were retained for testing in the next step.

6. Item testing

In the next phase of the scale development process, we tested the validated item set and collected data for further item reduction. Based on the recommendations of Boateng et al. (2018), which suggest a minimum sample size of 200 participants to ensure robust analysis, we conducted a study with 259 participants. The participants assessed seven scenarios involving novel technologies using a set of 58 items. These scenarios were selected and validated through a focus group to ensure their suitability as stimuli for the questionnaire. The scenarios used in this manuscript can be found in the supplementary material.

6.1. Focus group: Scenario validation

As there is no pre-defined way to induce people's expectations, we conducted a focus group with five participants to discuss the validity of the proposed seven scenarios and potential strategies to improve them. Participants could propose alterations to scenarios during the discussion. The goal of this step was to ensure that the selected set of scenarios is effectively suitable as a stimulus for eliciting user expectations.

Procedure Initially, participants received a brief introduction to the study and completed a demographic survey. Following this, we distributed printed copies of the seven scenarios to each participant. The group discussion was structured into two phases. In the first phase, participants individually reviewed each scenario, noting their thoughts and ranking the scenarios from most to least expectation-eliciting. In

Table 2
Overview of the backgrounds and expertise of the 11 experts who participated in the item validation. Including scientific degree and domain as well as self-assessed expertise in psychology (Psy.), HCI, System Development (Sys), and development (Dev.) of quantitative instruments.

	Gender	Age	Degree	Domain	Occupation	Psy. Exp.	HCI Exp.	Sys. Exp.	Dev. Exp.
1	Female	33	Doctoral	Res.	Researcher	■■■□□	■■■■■	■■■■□	■■□□□
2	Female	36	Doctoral	HCI	Professor	■■■■□	■■■■■	■■■■□	■■■■■
3	Male	35	Doctoral	HCI, AI	Professor	■■■□□	■■■■■	■■■■■	■■■■■
4	Male	27	Master's	HCI Res.	Researcher	■■■□□	■■■■■	■■■■□	■■■■■
5	Male	31	Master's	CS	Researcher	■■■□□	■■■■□	■■■□□	■■■□□
6	Male	37	Doctoral	HCI	Researcher	■■□□□	■■■■■	■■■■□	■■□□□
7	Female	33	Doctoral	Acad. Res.	Researcher	■■■□□	■■■■■	■■■■□	■■□□□
8	Male	29	Master's	HRI	Researcher	■■■□□	■■■■□	■■■□□	■■■■□
9	Female	34	Doctoral	HCI	Researcher	■■■■□	■■■■■	■■■■□	■■■■■
10	Female	34	Doctoral	HCI	Researcher	■■■■□	■■■■■	■■■■■	■■■■■
11	Male	35	Doctoral	CS, AI	Researcher	■□□□□	■■■■■	■■■■■	■■■■□

the second phase, we facilitated an open discussion where participants were encouraged to explain the reasoning behind their rankings. The facilitator ensured that all scenarios were thoroughly discussed. In the final stage, the group collaboratively agreed on a consensus ranking of the scenarios. Subsequently, participants were asked to propose modifications to the scenarios to enhance the range of their impact, aiming to increase the highest eliciting (positive expectations) scenarios while decreasing the impact of the lowest ones (negative expectations).

Participants Five participants (one female, four male) were recruited from the facilities of [Anonymized] through convenience sampling to participate in an in-person group discussion. Participation was voluntary, and participants were informed that they could withdraw at any time. The mean age of the participants was 23 years old (M = 23.60, SD = 2.33). The session lasted approximately 45 min, and participants did not receive any compensation for their involvement.

Results We collected individual participant rankings, the overall group ranking, and participants' comments from the printed articles. These results are summarized in Table 3. From the independent rankings, we observed that each scenario, except for the Nothing Earbuds, was ranked among the top three in terms of positive expectations by at least one participant. The Earbuds scenario also had the lowest standard deviation, indicating consensus among participants regarding its low expectation level. In contrast, the other scenarios exhibited a higher spread, reflecting a broader range of opinions. The Binoculars and Rabbit R1 devices elicited the highest expectations according to the rankings. This evaluation confirms that the selected scenarios effectively elicit a diverse range of user expectations and are suitable as stimuli for the current questionnaire.

6.2. Procedure

Participants were first requested to provide informed consent. Following consent, we collected demographic information before presenting them with one of seven scenarios described in the survey. To ensure that the respondents had no previous experience with the devices mentioned, we included a button labeled, "I have used, tested, or experienced this technology personally". If participants clicked this button, they were shown a different scenario. If a participant reported having experienced all the technologies in the scenarios, they were excluded from the sample.

Then, participants responded to three comprehension check questions and were then asked to explain the technology in one sentence. Subsequently, they rated 58 items on a Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree).

6.3. Participants

To collect user data for the factor analysis, we distributed a survey to participants from the USA and UK using Prolific[1] as the recruitment platform. We collected data from native English speakers, as the current scale is developed in English. We obtained responses from 325 participants, from which, after filtering for comprehension of the text, attention checks, and click latency analysis, we removed 67 participants. This led to a total of 259 valid responses from 126 females, 129 males, and four non-binary individuals. Participants had an average age of 37 years old ($M = 37.28, SD = 12.72$). The survey took approximately 10 min to complete ($M = 10.02, SD = 5.92$).

7. Item reduction

We analyzed the data obtained from the previous step to determine the optimal number of factors and reduce the number of items. This was done using Exploratory Factor Analysis (EFA), a method that reveals the underlying structure of the items by modeling the observed variables in relation to latent factors. The detailed procedure is described in the following section.

7.1. Item pre-processing and adequacy testing

For the item analysis, we first inverted the negatively worded items to ensure consistency in direction. We then examined the distribution of all items and removed those (n = 10) that showed high skewness or kurtosis, or that demonstrated poor item discrimination. To assess the dataset's suitability for factor analysis, we conducted the Kaiser–Meyer–Olkin (KMO) test for sampling adequacy. The resulting KMO

[1] https://prolific.com/.

Table 3
Categorization of scenarios, rater scores with average, Standard Deviation (SD), and overall ranking.

Scenario	Code	Innovation	Writing style	R1	R2	R3	R4	R5	Mean	SD	Rank
TinyPod	S1	Incremental	Neutral	● 3	● 1	● 2	❍ 6	❍ 7	3.8	2.32	3
Fiio DM13	S2	Incremental	Neutral	❍ 7	● 3	❍ 6	● 2	● 3	4.2	1.94	4
Sim-Lab Steering Wheel	S3	High	Heightening	● 1	❍ 5	❍ 7	❍ 7	❍ 4	4.8	2.23	5
Unistellar Binoculars	S4	High	Heightening	● 2	❍ 4	● 3	● 1	● 1	2.2	1.17	1
Rabbit R1	S5	Moderate	Heightening	❍ 6	● 2	● 1	❍ 4	● 2	3.0	1.79	2
Nothing's ChatGPT Earbuds	S6	Moderate	Neutral	❍ 4	❍ 6	❍ 4	❍ 5	❍ 6	5.0	0.89	7
Humane AI Pin	S7	High	Lowering	❍ 5	❍ 7	❍ 5	● 3	❍ 5	5.0	1.26	6

Table 4
The revised version of the iExpect scale consists of 22 items grouped in three factors: Anticipated Ease of Use (EU), Anticipated Usefulness (US), and Anticipated Enjoyment (AE), with item loadings reported. Loadings under 0.2 are omitted. Primary loadings used for scoring are in **Bold**.

Item	Code	Origin	Factor loadings		
			EU	US	AE
Learning how to use the technology will be easy for me	EU_1	UTAUT	**0.884**		
I am knowledgeable about how to use the technology	EU_2	TAM	**0.819**		−0.131
The operating procedures will be simple to understand	EU_3	meCUE 2.0	**0.779**		
It will be quickly apparent how to use the technology	EU_4	meCUE 2.0	**0.778**		
The technology will be easy to use	EU_5	UTAUT	**0.750**	0.150	
The technology will be confusing **(R)**	EU_6	UEQ	**0.742**		
I feel confident in my abilities to use the technology efficiently	EU_7	ACTA	**0.740**		0.108
I will be able to use the technology without the support of a technical person	EU_8	Positive SUS	**0.731**	−0.186	0.138
It will be easy to get the technology to do what I want it to	EU_9	ACTA	**0.637**	0.108	0.133
The technology will make it easier to do my daily activities	US_1	UTAUT		**0.903**	
The technology will increase my productivity	US_2	UTAUT		**0.885**	
The technology will increase the quality of my work output	US_3	UTAUT		**0.842**	
The technology will be useful in my daily life	US_4	UTAUT 2.0	0.109	**0.832**	
The technology will enable me to accomplish tasks more quickly	US_5	UTAUT		**0.826**	
The technology will help me reach my goals	US_6	meCUE 2.0		**0.821**	
The technology will increase the effectiveness of performing tasks	US_7	UTAUT		**0.688**	0.150
The technology will be enjoyable	AE_1	UEQ			**0.858**
The technology will be entertaining	AE_2	UTAUT 2.0			**0.860**
The technology will be fun to use	AE_3	ACTA			**0.860**
The technology will be boring **(R)**	AE_4	UEQ	−0.117		**0.689**
The technology will be stylish	AE_5	meCUE 2.0		0.131	**0.664**
The technology will make me happy	AE_6	meCUE 2.0		0.200	**0.621**

Measure of Sampling Adequacy (MSA) was 0.95, indicating excellent suitability (values closer to 1.0 are preferred). We also performed Bartlett's Test of Sphericity to test whether the correlation matrix significantly differs from an identity matrix. The test yielded a significant result, $\tilde{\chi}^2(1081) = 10933.87$, confirming that the variables are sufficiently correlated to justify factor analysis.

7.2. Exploratory factor analysis

We conducted an exploratory factor analysis (EFA) to uncover the underlying structure among the items (Streiner, 1994). To determine the optimal number of factors to retain, we used both parallel analysis (Horn, 1965) and scree plot inspection (Cattell & Vogelmann, 1977). The scree plot revealed a clear inflection point after the third factor, supporting a three-factor solution. This decision was further justified by the fact that each of the three retained factors had eigenvalues greater than 3.32.

We applied an oblimin (oblique) rotation to allow for correlations among factors, as recommended in the literature (Crawford, 1975). An orthogonal rotation was also tested but resulted in a poorer fit and failed to converge in the confirmatory factor analysis (CFA).[2] Oblimin rotation, by contrast, allows factors to correlate and aims to optimize factor loading structure by maximizing loading variance while minimizing cross-loadings (Crawford, 1975). Based on the rotated solution, we removed items with primary loadings below 0.40 or with substantial cross-loadings (i.e., above 0.20 on multiple factors; n = 3). We also removed three additional items due to extreme loadings (too high or too low), and merged two highly similar items in the final step of item refinement. The scale now encompassed 22 items distributed across three factors, with nine, seven, and six items, respectively. The model had a good fit, with KMO $MSA = 0.94$, Tucker Lewis Index of factoring reliability $TLI = 0.956$, and the Root Mean Square Error of Approximation $RMSEA = 0.056$. Table 4 presents the results of the exploratory factor analysis, namely the factors, items, and their corresponding loadings.

[2] Orthogonal rotation yielded an inferior fit and did not converge in the CFA.

Based on the content of each factor, we named the first factor Anticipated Ease of Use (EU), as it includes elements related to learning how to use the technology and the perceived ease of use. The second factor, named Anticipated Usefulness (US), encompasses items that focus on how well the technology supports the user's tasks. Finally, we labeled the third factor Anticipated Enjoyment (AE), as it includes items assessing the potential enjoyment derived from using the technology. The internal consistency of the scales, as measured by Cronbach's alpha, was high: $\alpha = .932$ for EU, $\alpha = .945$ for US, and $\alpha = .908$ for AE, suggesting good internal consistency (Cortina, 1993).

Dimension visualization: In Fig. 2, we show the three-factor solution scores for each scenario in a radar plot. Notably, scenarios 3 and 4 scored higher in Anticipated Usefulness, while scenarios 1 and 2 scored higher in Anticipated Ease of Use, scenarios 5, 6, and 7 had comparatively lower scores, with scenario 7 having the overall lower score.

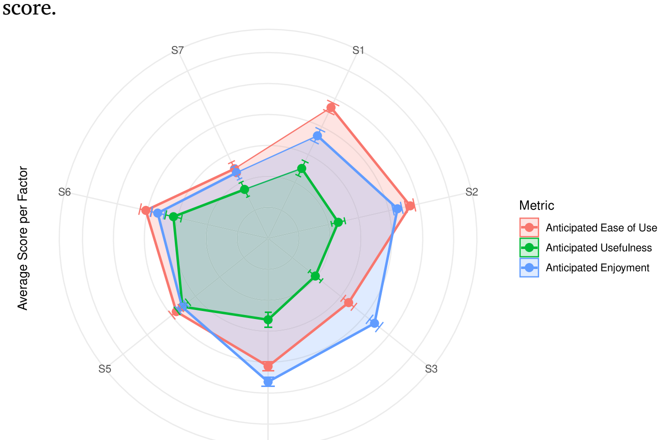

Fig. 2. Scenario scores for the resulting dimensions of the factor analysis.

8. Scale evaluation

In this step, we evaluated the obtained three-factor and 22-item solution found in the previous step across three assessments: (1) We first conducted a CFA to verify the factor structure identified in the previous step; (2) To assess the temporal stability of the scale, we performed a test–retest reliability analysis; (3) We then assessed discriminant, convergent and concurrent validity to ensure that the scale measures a construct distinct from related constructs, and can predict outcomes of connected concepts. A new sample (N = 278) was gathered to test the dimensionality, and a subset (N = 93) was recruited to participate again two weeks later for the temporal stability assessment.

8.1. Dimensionality test

To confirm the validity of the identified factor structure, it is relevant to verify whether the dimensions of the solution remain consistent (Anticipated Ease of Use, Anticipated Usefulness, Anticipated Enjoyment). Previously, we identified a three-factor solution with 22 items. Following Boateng et al. (2018), we conducted a CFA with a new participant sample ($N = 278$) to test if this structure holds.

8.2. Participants

To collect user data for the factor analysis, we distributed a survey to participants from the USA and UK using Prolific as the recruitment platform. We collected data from native English speakers, as the current scale is developed in English. We obtained responses from 332 participants, from which, after filtering for comprehension of the text, attention checks, and click latency analysis, we removed 54 participants. Our sample thus consists of 278 participants, including 140 females, 136 males, and 2 non-binary individuals. Participants had an average age of 37 years old ($M = 37.12, SD = 12.96$). The survey took approximately 4 min to complete ($M = 7.23, SD = 3.87$). The participants were compensated with 9 GBP per hour. All participants were informed that their participation was voluntary and that they could withdraw from the study at any time without consequence. They were also assured that all collected data would be anonymized before analysis.

Procedure Participants were first requested to provide informed consent. Following consent, we collected demographic information before presenting them with one of the seven scenarios described above. To ensure that the respondents had no previous experience with the devices mentioned, we included a button labeled, "I have used, tested, or experienced this technology personally". If participants clicked this button, they were shown a different scenario. If a participant reported having experienced all the technologies in the scenarios, they were excluded from the sample. After reading a scenario, participants responded to three comprehension check questions and were then asked to explain the technology in one sentence. Subsequently, they rated the 22 items on a Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree).

Confirmatory factor analysis In order to assess the validity of the iExpect scale's structure, we conducted a CFA. This statistical procedure allowed us to confirm the dimensionality of our proposed factor model. The solution had three intercorrelated factors; see again Table 4. We fitted the model with lavaan (Rosseel, 2012) using a maximum likelihood estimator with the NLMINB optimizer (56 iterations).

The confirmatory factor analysis indicated an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was 0.087, which is within an acceptable range. The Comparative Fit Index (CFI) was 0.937, and the Standardized Root Mean Square Residual (SRMR) was 0.077. Both CFI and SRMR values fall within commonly accepted thresholds for good fit (CFI $\geq$ 0.95 and SRMR $\leq$ 0.08). Additionally, internal consistency was high across all three factors: EU ($\alpha = .964$), US ($\alpha = .953$), and AE ($\alpha = .926$), indicating strong reliability. Together, these results support the validity of the proposed three-factor, 22-item solution.

8.3. Temporal stability

Temporal stability is an essential aspect of scale assessment, as it ensures validity for longitudinal studies and repeated measures, indicating that the instrument is not significantly influenced by external factors. Temporal stability refers to the consistency of a scale in producing similar results when administered to the same participants at different time points (Boateng et al., 2018). To evaluate the temporal stability of the iExpect solution, we conducted a test–retest reliability assessment. To assess test–retest reliability, a common step in scale development (e.g., Bentvelzen et al., 2021; Woźniak et al., 2021), we evaluated the temporal stability of the measure. We re-contacted 93 participants from the original CFA sample and collected their responses again two weeks later, allowing for comparison across the two time points.

Procedure The procedure was identical to the one described in the confirmatory factor analysis. The main difference from this procedure is that the participants were automatically assigned to the same scenario assigned in the first data collection. Therefore, participants were not allowed to change the scenario this time. Additionally, after participants completed the iExpect items, we distributed the Basic Psychological Needs Scale for Technology Use (BPN-TU) (Moradbakhti et al., 2024), the Big Five Inventory (BFI-10) (Rammstedt & John, 2007) and the Abbreviated Technology Anxiety Scale (ATAS) (Wilson et al., 2023) for discriminant and convergent validity.

Participants Approximately 33% of participants ($n = 93$) from the previous study were recruited again via Prolific for this follow-up data collection. The sample included 49 females and 44 males; none identified as non-binary or other, with a mean age of 39 years old ($M = 39.20, SD = 13.48$). The compensation and consent procedures remained consistent with those in the prior study. The survey was administered online and took respondents an average of 8 min to complete ($M = 8.53, SD = 4.99$).

Results To assess the reliability of the overall model, we computed the Intraclass Correlation Coefficient (ICC) using a two-way random effects model. We considered both consistency and agreement types of ICCs. Additionally, we calculated Spearman correlations for all the subscales and total scale. The consistency (ICC = 0.867, 95% CI = 0.806 to 0.91, $p < 0.001$) and the agreement (ICC = 0.867, 95% CI = 0.806 to 0.91, $p < 0.001$), and Spearman correlation ($\rho = .821$, $p < 0.001$) were found to be high for the overall scale suggesting a high degree of reliability in the overall scale items. Similarly, each of the three subscales yielded satisfactory consistency and agreement values; EU Consistency (ICC = .851, 95% CI = .783 to .899, $p < 0.001$), Agreement (ICC = .843, 95% CI = .767 to .895) and $\rho = .850$, $p < 0.001$, US Consistency (ICC = .747, 95% CI = .642 to .825, $p < 0.001$), Agreement (ICC = .749, 95% CI = .644 to .827, $p < 0.001$) and $\rho = .722$, $p < 0.001$, AE Consistency (ICC

Table 5
The two-way single-measurement intraclass correlation coefficients (ICC) calculated for Anticipated Ease of Use (EU), Anticipated Usefulness (US), and Anticipated Enjoyment (AE) factors of the Inventory of User Expectations for Technology (iExpect) scale.

Factor	iExpect-EU			iExpect-US			iExpect-AE		
	κ	F	p	κ	F	p	κ	F	p
Consistency	.851	$F(92,92) = 12.4$	<.005	.747	$F(92,92) = 6.91$	<.005	.863	$F(92,92) = 13.6$	<.005
Agreement	.843	$F(92,74.1) = 12.4$	<.005	.749	$F(92,92.1) = 6.91$	<.005	.862	$F(92,92.7) = 13.6$	<.005

= .863, 95% CI = .800 to .907, $p < 0.001$), Agreement (ICC = .862, 95% CI = .799 to .906, $p < 0.001$) and ρ = .837, $p < 0.001$. Table 5 shows a summary of the results per factor.

To visualize the reliability of the iExpect scale, we employed the Bland-Altman method (Bland & Altman, 1999). We used Bland–Altman plots to visualize the agreement between test and retest scores. For each participant, we plotted the mean difference between the two sessions against the average of their scores. The dashed horizontal lines indicate the limits of agreement, representing the 95% confidence interval around the mean difference (Bland & Altman, 1999; Welsch et al., 2021). These limits define the range within which 95% of the differences between test and retest scores are expected to lie.

In the plot (see Fig. 3), the mean difference near zero (dotted line) indicates that the iExpect scale demonstrates absolute temporal stability on average. The distribution around zero further suggests that reliability is not influenced by the mean score. This supports the scale's suitability for administration at different time points, making it appropriate for use in both between-groups and repeated-measures designs.

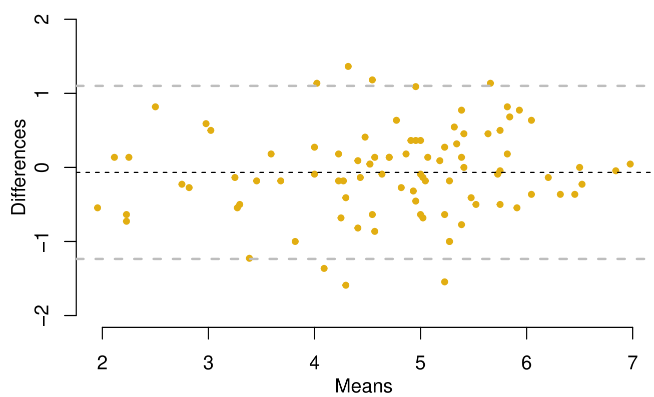

Fig. 3. Bland Altman plot showing the iExpect Mean score for both test sessions in each participant as a function of the difference between both testings. There is no indication of skew in the data. Therefore, the small error found seems to vary independently of the Mean score.

8.4. Construct validity

Following the validation of the scale's factor structure and temporal stability, one of the last steps involves assessing the construct validity, as recommended by Boateng et al. (2018). In this section, we evaluate the construct validity of iExpect across three dimensions: (1) Discriminant validity, to ensure that the construct is distinct and independent from related constructs; (2) Convergent validity, to confirm that the construct correlates with similar concepts; and (3) Concurrent validity, to demonstrate that the construct can predict the outcomes of associated concepts.

Measures: To test discriminant and convergent validity of the iExpect scale, we selected three instruments available in the literature, including all their subscales; these instruments are listed below:

BPN Basic Psychological Needs Scale for Technology Use (BPN-TU) (Moradbakhti et al., 2024): The BPN-TU scale is a psychometric tool to assess how well the use of technology satisfies three core psychological needs: **autonomy** (BPN_{Aut}), **competence** (BPN_{Com}), and **relatedness** (to others-BPN_{Rel_O}-, and to the technology -BPN_{Rel_T}-). It is grounded in Self-Determination Theory and evaluates users' experiences with technology by measuring how these interactions fulfill their intrinsic motivations.

BFI Big Five Inventory (BFI-10) (Rammstedt & John, 2007): is used to measure five major dimensions of personality: **Extraversion** (BFI_{Extra}), **Conscientiousness** (BFI_{Consc}), **Neuroticism** (BFI_{Neuro}), **Openness** (BFI_{Open}), and **Agreeableness** (BFI_{Agre}). It provides insights into individual personality traits and is widely utilized in psychological research.

ATAS Abbreviated Technology Anxiety Scale (ATAS) (Wilson et al., 2023): is an **unidimensional** tool to measure individual's anxiety or discomfort related to the use of technology. ATAS is particularly useful in contexts where understanding the psychological barriers to technology adoption and use is critical.

We hypothesize that the BPN scale, particularly the BPN_{Aut} subscale, will positively correlate with the iExpect-EU. This is because the BPN_{Aut} subscale reflects the user's sense of control over the technology and their ability to use it predictably. Similarly, we expect a positive correlation between the BPN_{Com} subscale and the iExpect-US, due to the alignment with the goal-oriented anticipation measured by US.

While personality traits, as measured by the BFI-10 inventory, may influence attitudes toward novel technologies, we anticipate that the iExpect construct will be largely uncorrelated with these traits. The iExpect is specifically focused on the technology itself and the anticipation of its use, which we believe remains independent of personality traits, particularly in the pre-use stage when users lack experience with the technology. However, it is plausible that certain factors, such as BFI_{Agre}, which may involve elements of traits like trust, could show a slight correlation. This is because trust in the potential benefits of the technology may play a role in shaping expectations.

Finally, we hypothesize that the $ATAS$ scale will show a negative correlation with our construct. This expectation is based on the idea that the anticipation of benefits from using technology is inversely related to the anxiety that using technology might provoke.

Participants: The construct validity assessments were made in the same subset of participants from the temporal stability assessment.

Internal consistency The split-half reliability of the iExpect was evaluated using the split-half function with 1000 samples, demonstrating strong internal consistency across multiple metrics. The maximum split-half reliability (λ_4) was 0.99. Both the average split-half reliability ($\bar{\lambda}$) and Cronbach's (α) were 0.96. The minimum split-half reliability (β) was observed at 0.74. We also observed similar values for iExpect-EU (λ_4 = .97, α = .96, $\bar{\lambda}$ = .96, β = .93), iExpect-US (λ_4 = .97, α = .96, $\bar{\lambda}$ = .96, β = .95), and iExpect-AE (λ_4 = .95, α = .93, $\bar{\lambda}$ = .93, β = .91) indicating high internal consistency and suggesting that the items within the iExpect are reliably measuring the intended construct.

Discriminant validity To ensure the independence of our construct from related scales, we followed the method proposed by Rönkkö and Cho (2022). We first computed models for the iExpect scale along with the previously mentioned instruments. Subsequently, for the $ATAS$ scale, we observed a negative correlation with both the iExpect-EU and iExpect-AE subscales, aligning with the anticipated relationships based on the underlying concepts of each construct. As for the BFI-10 subscales — BFI_{Extra}, BFI_{Consc}, BFI_{Neuro}, BFI_{Open}, and BFI_{Agre} — each showed small, negligible correlations with the proposed constructs, in fact, all the upper confidence intervals (CI_{Upper}) of the correlations where lower than 0.68, Rönkkö and Cho (2022) suggest discriminant validity when $CI_{\text{Upper}} < 0.8$. Table 6 shows the computations for the discriminant validity instruments. Given these results, it is possible to assert that the iExpect scale factors are independent and distinct from connected instruments and concepts.

Convergent validity For the convergent validity assessment, we applied the same procedure as we did for the discriminant validity. We observed correlations across all subscales for the instruments used to assess convergent validity. Specifically, BPN_{Aut} and BPN_{Com} showed positive correlations with the three factors of the iExpect scale. BPN_{Rel_O} and BPN_{Rel_T} demonstrated weak correlations with iExpect-EU but had stronger correlations with iExpect-US and iExpect-AE, particularly with the latter. Table 7 provides a summary of the convergent validity results. Additionally, following Cheung et al. (2024)'s

Table 6
Discriminant validity calculations for the ATAS questionnaire, the BFI-10 subscales and the iExpect factors.

Scale	iExpect-EU				iExpect-US				iExpect-AE			
	$Corr$	df	χ^2	$RMSEA$	$Corr$	df	χ^2	$RMSEA$	$Corr$	df	χ^2	$RMSEA$
$ATAS$	−0.49	459	820.56	0.39	−0.01	459	876.33	0.87	−0.33	459	837.47	0.58
BFI_{Extr}	−0.08	302	661.62	0.82	−0.10	302	660.04	0.81	0.01	302	672.38	0.89
BFI_{Cons}	0.10	302	660.04	0.81	0.04	302	667.79	0.86	0.10	302	659.92	0.81
BFI_{Neur}	−0.24	302	641.63	0.68	−0.03	302	668.33	0.87	−0.22	302	644.39	0.70
BFI_{Open}	0.01	302	671.93	0.89	−0.11	302	658.00	0.80	−0.15	302	653.10	0.70
BFI_{Agre}	0.32	302	632.71	0.61	0.26	302	639.82	0.67	0.35	302	629.07	0.57

recommendations, we analyzed the following criteria: (1) The ω-values > 0.7, (2) the standardized factor loadings > 0.4, and (3) the AVE values are > 0.5. All the scales presented an ω higher than 0.7, similarly the AVE was also higher than 0.5. Regarding the standardized factor loadings, we found that BPN_{Aut} and BPN_{Com} presented a high value for all three factors, yet BPN_{Rel_O} presented a high value for US and AE mostly, while BPN_{Rel_T} presented a high value only for AE. In all, considering the redundant number of tests, the result evidences a convergent validity, highlighting that the proposed scale aligns with conceptually related instruments.

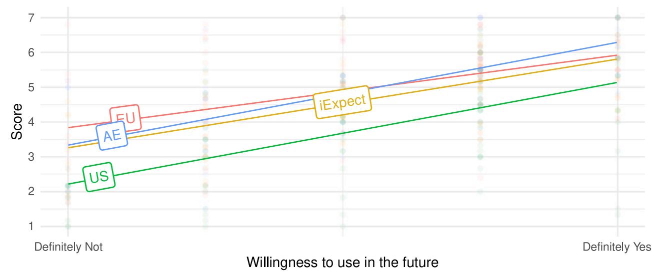

Fig. 4. Concurrent validity of the iExpect scale; Willingness to use a technology in the future as predicted by the iExpect scale and each of its factors (Anticipated Ease of Use (EU), Anticipated Usefulness (US), and, Anticipated Enjoyment (AE)). The results show that at a higher score of the iExpect, there is a higher willingness to use a given technology in the future as measured by the responses of the study participants.

Concurrent validity To evaluate the concurrent validity of the iExpect, we examined how effectively the scale and its factors predict the willingness to use technology in the future, Fig. 4. A multiple linear regression analysis was conducted to assess the impact of three independent variables (EU, US, and AE) on participants' responses to the question, *"Would you be willing to use this product in the future?"*. The regression model was statistically significant, $F(3, 89) = 37.887$, $p < 0.001$, indicating that the independent variables collectively accounted for a significant portion of the variance in the dependent variable.

9. Discussion

In this paper, we present the development and validation of the 22-item iExpect scale, designed to measure users' expectations of a technology prior to interaction. Through factor analysis, we identified a three-factor structure: EU, which reflects anticipated ease of use; US, which captures anticipated usefulness; and AE, which measures anticipated enjoyment of using the interactive system.

9.1. Scale scoring and independent use of the subscales

The iExpect scale is scored on a seven-point Likert scale from Strongly Disagree (1) to Strongly Agree (7). Items EU-6 and AE-4 are reverse-scored. Higher scores indicate higher user expectations about a given technology:

In the full scoring system of the iExpect scale, it is advisable to calculate the arithmetic mean of all the items to obtain the overall iExpect score. This scoring provides an overview of user expectations about a technology.

$$iExpect = \frac{EU + US + AE}{3} \tag{1}$$

Yet, we also recommend analyzing the individual factors of the scale given their individual relevance and the length of the overall questionnaire, so if a researcher would need to look at the specific insight of AE for example, distributing and computing only the elements of this subscale would be possible. In order to do this, we suggest first reversing the relevant items for EU and AE and then calculating the arithmetic mean of the items of the factor. The interpretation of the independent subscales is similar to the overall scale as they have the same valence: higher scores indicate higher expectations about a given technology.

$$EU = (EU_1 + EU_2 + EU_3 + EU_4 + EU_5 + EU_{6_R} + EU_7 + EU_8 + EU_9)/9$$
$$US = (US_1 + US_2 + US_3 + US_4 + US_5 + US_6 + US_7)/7$$
$$AE = (AE_1 + AE_2 + AE_3 + AE_{4_R} + AE_5 + AE_6)/6$$

9.2. Scale structure and content

The iExpect sub-scales were validated by comparing different scenarios and in confirmatory factor analysis, showing a good fit based on several fit indices, excellent internal consistency, and very good test–retest reliability. We have evaluated the validity of iExpect across studies. In the Expert Feedback section, we show experts' agreement regarding our tested items for the iExpect and, therefore, the validity of the content of our item pool is substantiated. Furthermore, our three-factor structure aligns with the TAM model of technology acceptance. The TAM model, originally developed by Davis (1985), posits that user acceptance of technology is primarily determined by two key factors: Perceived Usefulness and Perceived Ease of Use. Our three-factor structure, encompassing Anticipated Ease of Use, Anticipated Usefulness and Anticipated Enjoyment, not only aligns with these foundational constructs but also extends the model by incorporating an affective

Table 7
Convergent validity calculations for the BPN subscales and the iExpect factors, SFL represents the Standardized Factor Loading, (which resembles the correlation between the factors).

Scale	ω	AVE	iExpect-EU			iExpect-US			iExpect-AE		
			SFL	df	χ^2	SFL	df	χ^2	SFL	df	χ^2
BPN_{Aut}	0.90	0.77	0.70	507	983.77	0.57	507	990.75	0.79	507	982.83
BPN_{Com}	0.87	0.70	0.59	507	988.79	0.67	507	984.21	0.78	507	982.83
BPN_{Rel_O}	0.88	0.68	0.24	507	1020.96	0.51	507	994.73	0.43	507	1001.48
BPN_{Rel_T}	0.83	0.62	0.33	507	1009.56	0.39	507	1004.71	0.48	507	996.71

dimension (Anticipated Enjoyment) which has been increasingly recognized as an important factor in user experience and technology acceptance in subsequent extensions of TAM (Bargas-Avila & Hornbæk, 2011; Davis, 1985; Kujala & Miron-Shatz, 2015; Kujala et al., 2011; Thong et al., 2011; van Schaik & Ling, 2011), such as UTAUT2 (Hedonic Motivation) (Madigan et al., 2017; Maillet et al., 2015; Venkatesh et al., 2003, 2012). The Anticipated Enjoyment factor specifically draws from established constructs in user experience research, including Hedonic Motivation from UTAUT2 (Venkatesh et al., 2012), Attractiveness from the UEQ (Schrepp et al., 2017), Intrinsic Self-Regulation from ACTA (Peters et al., 2018), and Positive Emotions and Aesthetics from meCUE 2.0 (Minge et al., 2017). This dimension captures users' anticipatory affective responses — expectations of pleasure, aesthetic appeal, and emotional satisfaction — that form before interaction with the technology, distinct from cognitive evaluations of usefulness or ease of use.

Aside from content validity, we could also establish concurrent validity. The factors of the scale could predict willingness to use technology in the future, demonstrating the criterion validity of our scales. Therefore, we deem our scale to be practically useful for capturing aspects of the intention to use the technology before any interaction, which sets it apart from the TAM-based questionnaires and scales that focus on acceptance and intention to use technology once participants have interacted with it (Davis, 1985; Madigan et al., 2017; Maillet et al., 2015; Venkatesh et al., 2003, 2012). Interestingly, in the final set of items of iExpect, there is no reference to social aspects of technology use, as covered in the meCue scale (Minge et al., 2017). and our adaption, hinting at the possibility that before interaction, the social acceptability aspects of technology expectations are not as clear. While this view aligns with technological determinism (see Wyatt (2008)), social change coming from technological adoption, social acceptability might still be a key part of not yet-used technologies if the product vision is defined and in public discourse such as in social robotics (Syrdal et al., 2009) or human augmentation (Villa, Niess et al., 2023).

We found a distinct pattern regarding the correlation to related and unrelated measures, showing convergent and discriminant validity. While we found no correlations for most BFI-10 measures and a negative correlation to the ATAS, we found that our subscales correlated with the BPN scale. Given that both the ATAS and the BPN scale relate to emotional and motivational aspects regarding the anticipation of technology use, we can speculate that expectations measured with the iExpect might link the distal emotional and motivational aspects to expectations of interaction, which are relatively proximal, i.e., map well onto the willingness to use. Moreover, while our measure was highly reliable, with very few relative and absolute changes across time, given this link to emotional and motivational aspects of technology use, the measure should be sensitive to dynamic changes in expectations across time if new information about technology is given to potential users (e.g., from sneak-peak of a product to the presentation keynote).

9.3. Implications

Our research has three sets of implications for practice, theory, and methods. First, with our scale at hand, practitioners can now evaluate how descriptions of different technologies, even in the prototyping stage, can elicit expectations. In product development, our scale could be used to understand aspects of purchase intentions for interactive technologies and map out which aspects are relevant to different user groups and demographics. However, it is important to note that in this regard, further research is needed to understand individual differences in expectations toward interactive technologies.

Second, with the iExpect, we have a new method. We now have a tool for measuring expectations regarding technologies before interaction. Given that scales like the ones used (Brown et al., 2014, 2008) have been used to anticipate technology use before interaction in practice but were constructed for acceptance after interaction and use of technology, we can resolve a long-standing mismatch in the TAM literature and provide now a new measure that is designed to measure expectations of technology.

Third, on a larger scale, the iExpect could be a new tool for evaluation efforts as expectation-confirmation / disconfirmation (Brown et al., 2014), as well as placebo effects (Denisova & Cairns, 2015; Kloft et al., 2024; Kosch et al., 2022; Villa, Niess et al., 2023), can undermine evaluation efforts. By capturing expectations upfront, the iExpect allows researchers and UX practitioners to anticipate biases (due to very high or very low expectations) that can skew evaluation results, ensuring that assessments reflect the technology's true performance, usability, and user experience rather than being influenced by preconceived notions or expectations. Nevertheless, it is crucial to investigate whether these expectations stem from unrealistic perceptions, misunderstandings, exaggerated marketing, or genuine beliefs. To address this, researchers and practitioners should interview users to understand the reasons behind their expectations, ensuring that subsequent guidance or product adjustments are based on whether expectations are well-founded or if there is a misalignment between user perception and the technology's actual capabilities.

9.4. Core contributions

This research offers three primary contributions: (1) First, at a theoretical level, it refines expectation-related constructs in HCI by clearly distinguishing pre-use anticipatory beliefs from post-use perceptions in TAM- and UTAUT-based frameworks, thereby addressing a temporal gap in technology acceptance theory. In particular, Anticipated Enjoyment (AE) clarifies the role of affective expectations as a pre-interaction construct that may bias later experiential evaluations. (2) Second, at a methodological level, we provide the first psychometrically validated instrument specifically designed to measure expectations prior to interaction, resolving a long-standing mismatch in the use of post-use acceptance scales for pre-use assessment. (3) Third, at a practical level, the iExpect enables researchers and practitioners to systematically control for expectation-driven biases in product evaluation, communication, and early-stage deployment of emerging technologies.

9.5. Limitations

Despite the strengths of our study, there are limitations that should be acknowledged. First, our research did not include a longitudinal component, rendering us unable to capture dynamic changes (for a dynamic retrospective UX curve, see Kujala et al. (2011)) in expectations over time or reasons for the changing expectations. Furthermore, we did not study how user expectations influence user experience or how this influence evolves over use time such as in Kujala et al. (2017). This limits our understanding of how expectations evolve over time and how they might influence long-term user behavior and satisfaction.

Second, our scale was only tested in English-speaking populations. While this allows for a focused analysis, it also limits the generalizability of our findings to non-English-speaking contexts. Future research should consider adapting and validating the iExpect in other languages and cultural contexts to ensure broader applicability (for a recent AI-based tool to develop translations for research scales, see Haavisto and Welsch (2024)). Third, although our scale is designed to be more concise than the UTAUT model (Madigan et al., 2017; Maillet et al., 2015; Venkatesh et al., 2003, 2012) and other similar frameworks, it is still relatively long. However, it is important to note that the scale is more economically focused, specifically targeting pre-interaction expectations, unlike UTAUT, which emphasizes post-interaction experiences.

Fourth, while the iExpect shows promise as a tool for capturing expectations, its potential role as a marker for placebo effects or expectation confirmation/disconfirmation effects, e.g., as in Brown et al.

(2014) or Raita and Oulasvirta (2011), has yet to be empirically tested. This is a critical area for future research, as understanding these dynamics could significantly enhance the utility of iExpect in both practical and theoretical applications.

Fifth, while the iExpect provides a quantitative measure of user expectations, it does not capture the reasons behind those expectations. Users' beliefs, concerns, or specific contextual factors (e.g., in healthcare, the changing relationship between health professionals and patients (Kujala et al., 2018)) may influence their expectations. Without understanding the rationale behind the expectations, there is a risk of missing important factors that shape user experience and future product acceptance (for example, in healthcare, see Ludwick and Doucette (2009)). Future studies should complement the scale with qualitative methods, such as interviews, to uncover these underlying reasons and provide a more comprehensive understanding of user expectations.

10. Conclusion

We developed and validated the iExpect, a 22-item tool for measuring user expectations of interactive technologies across three dimensions: Anticipated Ease of Use, Anticipated Usefulness, and Anticipated Enjoyment. The scale proved reliable and valid, offering a theoretically grounded and methodologically robust instrument for assessing pre-use expectations. This work supplements existing acceptance models, introduces a dedicated pre-interaction measurement tool, and provides practical value for controlling expectation-driven biases in technology evaluation and deployment. Future applications can include its use in product evaluation and development cycles to better align design, communication, and research with user expectations.

CRediT authorship contribution statement

Steeven Villa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Kosch:** Writing – review & editing, Supervision. **Agnes Mercedes Kloft:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Jasper Quinn:** Investigation, Data curation. **Sari Kujala:** Writing – review & editing, Supervision, Methodology. **Robin Welsch:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Ethics statement

This research was conducted in accordance with the Finnish National Board on Research Integrity (TENK) guidelines and Aalto University's ethical review procedures. According to an internal ethical assessment and institutional policy, formal review by an institutional ethics committee was not required. The study involved minimal risk, included non-invasive methods such as online surveys and moderated focus group discussions, and did not involve the collection of sensitive personal data. All participants were informed about the study's purpose and procedures, assured of their right to withdraw at any time without consequence, and gave informed consent prior to participation. Data were anonymized before analysis to ensure participant confidentiality.

Declaration of Generative AI and AI-assisted technologies in the writing process

We recognize that author positionality shapes the perspectives and interpretations in this paper (Olteanu et al., 2023). We are researchers situated in Germany and Finland, with backgrounds in Cognitive Science, HCI, and Engineering. Our diverse perspectives informed our research framing and analysis. We used LLMs (GPT-5, Claude Sonnet 4) to receive feedback on text clarity. All analytical decisions and interpretive claims, specifically all interpretation, reflection, and discussion, were human-only. We disclose our AI usage (cf., El Ali et al. (2024)) in efforts to ensure transparency in AI-mediated scholarship (Elagroudy et al., 2024).

Funding

This work was conducted as part of the AI-Twin project, which is funded by the European Research Council (ERC-2024-ADG) as part of the European Union's Horizon 2020 research and innovation program (grant agreement no. 101200584).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chbr.2026.101014.

Data availability

Data will be made available on request.

References

Anderson, R. E. (1973). Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research*, *10*(1), 38–44. http://dx.doi.org/10.1177/002224377301000106, Publisher: SAGE Publications Inc.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. http://dx.doi.org/10.1080/10447310802205776, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447310802205776.

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2689–2698). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1978942.1979336.

Bentvelzen, M., Niess, J., Woźniak, M. a. P., & Woźniak, P. W. (2021). The development and validation of the technology-supported reflection inventory. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–8). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3411764.3445673, URL: https://dl.acm.org/doi/10.1145/3411764.3445673.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160.

Bly, S., Schilit, B., McDonald, D. W., Rosario, B., & Saint-Hilaire, Y. (2006). Broken expectations in the digital home. In *CHI EA '06, CHI '06 extended abstracts on human factors in computing systems* (pp. 568–573). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1125451.1125571.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. http://dx.doi.org/10.3389/fpubh.2018.00149, URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6004510/.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*(4), 445–454. http://dx.doi.org/10.1177/1745691613491271, Publisher: SAGE Publications Inc.

Bosch, E., Welsch, R., Ayach, T., Katins, C., & Kosch, T. (2024). The illusion of performance: The effect of phantom display refresh rates on user expectations and reaction times. In *CHI EA '24, Extended abstracts of the 2024 CHI conference on human factors in computing systems* (pp. 1–6). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3613905.3650875.

Brown, S. A., Venkatesh, V., & Goyal, S. (2014). Expectation confirmation in information systems research: A test of six competing models. *MIS Quarterly*, *38*(3), 729–A9. http://dx.doi.org/10.25300/MISQ/2014/38.3.05, URL: https://www-jstor-org.emedien.ub.uni-muenchen.de/stable/26634990 Publisher: Management Information Systems Research Center, University of Minnesota.

Brown, S. A., Venkatesh, V., Kuruzovich, J., & Massey, A. P. (2008). Expectation confirmation: An examination of three competing models. *Organizational Behavior and Human Decision Processes*, *105*(1), 52–66. http://dx.doi.org/10.1016/j.obhdp.2006.09.008, URL: https://www.sciencedirect.com/science/article/pii/S0749597806000914.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and kg criteria for determining the number of factors.. *Multivariate Behavioral Research*, *12 3*, 289–325.

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*(2), 745–783. http://dx.doi.org/10.1007/s10490-023-09871-y.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications.. *Journal of Applied Psychology*, *78*(1), 98.

Crawford, C. (1975). A comparison of the direct oblimin and primary parsimony methods of oblique rotation. *British Journal of Mathematical and Statistical Psychology*, *28*(2), 201–213.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems : theory and results* (Ph.D. thesis), Massachusetts Institute of Technology, URL: https://dspace.mit.edu/handle/1721.1/15192 Accepted: 2005-08-08T16:11:32Z.

den Ouden, E., Yuan, L., Sonnemans, P. J. M., & Brombacher, A. C. (2006). Quality and reliability problems from a consumer's perspective: an increasing problem overlooked by businesses? *Quality and Reliability Engineering International*, *22*(7), 821–838. http://dx.doi.org/10.1002/qre.766, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.766 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.766.

Denisova, A., & Cairns, P. (2015). The placebo effect in digital games: Phantom perception of adaptive artificial intelligence. In *Proceedings of the 2015 annual symposium on computer-human interaction in play* (pp. 23–33). London United Kingdom: ACM, http://dx.doi.org/10.1145/2793107.2793109.

Denisova, A., & Cairns, P. (2019). Player experience and deceptive expectations of difficulty adaptation in digital games. *Entertainment Computing*, *29*, 56–68. http://dx.doi.org/10.1016/j.entcom.2018.12.001, URL: https://www.sciencedirect.com/science/article/pii/S1875952118300521.

El Ali, A., Venkatraj, K. P., Morosoli, S., Naudts, L., Helberger, N., & Cesar, P. (2024). Transparent AI disclosure obligations: Who, what, when, where, why, how. In *CHI EA '24, Extended abstracts of the CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3613905.3650750.

Elagroudy, P., Li, J., Väänänen, K., Lukowicz, P., Ishii, H., Mackay, W. E., Churchill, E. F., Peters, A., Oulasvirta, A., Prada, R., Diening, A., Barbareschi, G., Gruenerbl, A., Kawaguchi, M., El Ali, A., Draxler, F., Welsch, R., & Schmidt, A. (2024). Transforming HCI research cycles using generative AI and "large whatever models" (LWMs). In *CHI EA '24, Extended abstracts of the CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3613905.3643977.

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–107. http://dx.doi.org/10.1038/scientificamerican1062-93, Place: US Publisher: Scientific American, Inc..

Haavisto, O., & Welsch, R. (2024). Questionnaires for everyone: Streamlining cross-cultural questionnaire adaptation with GPT-based translation quality evaluation. arXiv preprint arXiv:2407.20608.

Hartmann, J., De Angeli, A., & Sutcliffe, A. (2008). Framing the user experience: information biases on website quality judgement. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 855–864). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1357054.1357190.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

Irving, P. G., & Meyer, J. P. (1994). Reexamination of the met-expectations hypothesis: A longitudinal analysis. *Journal of Applied Psychology*, *79*(6), 937–949. http://dx.doi.org/10.1037/0021-9010.79.6.937, Place: US Publisher: American Psychological Association.

Klein, J. G. (1999). Developing negatives: Expectancy assimilation and contrast in product judgments. | advances in consumer research | ebscohost. URL: https://openurl.ebsco.com/contentitem/gcd:83144365?sid=ebsco:plink:crawler&id=ebsco:gcd:83144365 ISSN: 0098-9258 Issue: 1 Pages: 463 Volume: 26.

Kloft, A. M., Welsch, R., Kosch, T., & Villa, S. (2024). "AI enhances our performance, I have no doubt this one will do the same": The placebo effect is robust to negative descriptions of AI. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–24). Honolulu HI USA: ACM, http://dx.doi.org/10.1145/3613904.3642633, URL: https://dl.acm.org/doi/10.1145/3613904.3642633.

Kosch, T., Welsch, R., Chuang, L., & Schmidt, A. (2022). The placebo effect of artificial intelligence in human–computer interaction. *ACM Transactions on Computer-Human Interaction*, *29*(6), 1–32. http://dx.doi.org/10.1145/3529225, URL: https://dl.acm.org/doi/10.1145/3529225.

Kujala, S., Hörhammer, I., Kaipio, J., & Heponiemi, T. (2018). Health professionals' expectations of a national patient portal for self-management. *International Journal of Medical Informatics*, *117*, 82–87.

Kujala, S., & Miron-Shatz, T. (2015). The evolving role of expectations in long-term user experience. In *Proceedings of the 19th international academic mindtrek conference* (pp. 167–174). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2818187.2818271.

Kujala, S., Mugge, R., & Miron-Shatz, T. (2017). The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. *International Journal of Human-Computer Studies*, *98*, 51–61. http://dx.doi.org/10.1016/j.ijhcs.2016.09.011, URL: https://www.sciencedirect.com/science/article/pii/S1071581916301239.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX curve: A method for evaluating long-term user experience. *Interacting with Computers*, *23*(5), 473–483. http://dx.doi.org/10.1016/j.intcom.2011.06.005.

Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. http://dx.doi.org/10.1080/10447318.2018.1455307, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2018.1455307.

Ludwick, D. A., & Doucette, J. (2009). Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. *International Journal of Medical Informatics*, *78*(1), 22–31.

Madigan, R., Louw, T., Wilbrink, M., Schieben, A., & Merat, N. (2017). What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 55–64. http://dx.doi.org/10.1016/j.trf.2017.07.007, URL: https://www.sciencedirect.com/science/article/pii/S1369847816301620.

Maillet, E., Mathieu, L., & Sicotte, C. (2015). Modeling factors explaining the acceptance, actual use and satisfaction of nurses using an electronic patient record in acute care settings: An extension of the UTAUT. *International Journal of Medical Informatics*, *84*(1), 36–47. http://dx.doi.org/10.1016/j.ijmedinf.2014.09.004, URL: https://www.sciencedirect.com/science/article/pii/S1386505614001774.

Miceli, M. P. (1987). Effects of realistic job previews on newcomer affect and behavior. *Journal of Organizational Behavior Management*, http://dx.doi.org/10.1300/J075v08n01_06, URL: https://www.tandfonline.com/doi/abs/10.1300/J075v08n01_06 Publisher: Taylor & Francis Group.

Michalco, J., Simonsen, J. G., & Hornbæk, K. (2015). An exploration of the relation between expectations and user experience. *International Journal of Human–Computer Interaction*, *31*(9), 603–617. http://dx.doi.org/10.1080/10447318.2015.1065696, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2015.1065696.

Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The mecue questionnaire: A modular tool for measuring user experience. In M. Soares, C. Falcão, & T. Z. Ahram (Eds.), *Advances in ergonomics modeling, usability & special populations* (pp. 115–128). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-41685-4_11.

Moradbakhti, L., Leichtmann, B., & Mara, M. (2024). Development and validation of a basic psychological needs scale for technology use. *Psychological Test Adaptation and Development*, *5*(1), 26–45. http://dx.doi.org/10.1027/2698-1866/a000062, URL: https://econtent.hogrefe.com/doi/10.1027/2698-1866/a000062 Publisher: Hogrefe Publishing.

Olson, J. C., & Dover, P. A. (1979). Disconfirmation of consumer expectations through product trial. *Journal of Applied Psychology*, *64*(2), 179–189. http://dx.doi.org/10.1037/0021-9010.64.2.179, Place: US Publisher: American Psychological Association.

Olsson, T., & Salo, M. (2012). Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2779–2788). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2207676.2208677.

Olteanu, A., Ekstrand, M., Castillo, C., & Suh, J. (2023). Responsible AI research needs impact statements too. arXiv:2311.11776 URL: https://arxiv.org/abs/2311.11776.

Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence*, *5*(10), 1076–1086. http://dx.doi.org/10.1038/s42256-023-00720-7, URL: https://www.nature.com/articles/s42256-023-00720-7 Publisher: Nature Publishing Group.

Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology*, *9*, http://dx.doi.org/10.3389/fpsyg.2018.00797, URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00797/full Publisher: Frontiers.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. http://dx.doi.org/10.1002/nur.20147, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nur.20147 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nur.20147.

Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin*, *80*(2), 151–176. http://dx.doi.org/10.1037/h0034829, Place: US Publisher: American Psychological Association.

Pulakos, E. D., & Schmitt, N. (1983). A longitudinal study of a valence model approach for the prediction of job satisfaction of new employees. *Journal of Applied Psychology*, *68*(2), 307–312. http://dx.doi.org/10.1037/0021-9010.68.2.307, Place: US Publisher: American Psychological Association.

Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, *23*(4), 363–371. http://dx.doi.org/10.1016/j.intcom.2011.04.002, arXiv:https://academic.oup.com/iwc/article-pdf/23/4/363/2208237/iwc23-0363.pdf.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, *41*(1), 203–212. http://dx.doi.org/10.1016/j.jrp.2006.02.001, URL: https://www.sciencedirect.com/science/article/pii/S0092656606000195.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. http://dx.doi.org/10.1177/1094428120968614, Publisher: SAGE Publications Inc.

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Schmidmaier, M., Rupp, J., Cvetanova, D., & Mayer, S. (2024). Perceived empathy of technology scale (PETS): Measuring empathy of systems toward the user. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–18). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3613904.3642035, URL: https://dl.acm.org/doi/10.1145/3613904.3642035.

Schrepp, M., Thomaschewski, J., & Hinderks, A. (2017). Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(Regular Issue), 40–44. http://dx.doi.org/10.9781/ijimai.2017.445, URL: https://www.ijimai.org/journal/bibcite/reference/2604.

Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, *39*, 135–140.

Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study.

Thong, J. Y. L., Venkatesh, V., Xu, X., Hong, S.-J., & Tam, K. Y. (2011). Consumer acceptance of personal information and communication technology services. *IEEE Transactions on Engineering Management*, *58*(4), 613–625. http://dx.doi.org/10.1109/TEM.2010.2058851, URL: https://ieeexplore.ieee.org/abstract/document/5942151 Conference Name: IEEE Transactions on Engineering Management.

van Schaik, P., & Ling, J. (2008). Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with Computers*, *20*(3), 419–432. http://dx.doi.org/10.1016/j.intcom.2008.03.001.

van Schaik, P., & Ling, J. (2011). An integrated model of interaction experience for information retrieval in a web-based encyclopaedia. *Interacting with Computers*, *23*(1), 18–32. http://dx.doi.org/10.1016/j.intcom.2010.07.002.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. http://dx.doi.org/10.2307/30036540, URL: https://www.jstor.org/stable/30036540 Publisher: Management Information Systems Research Center, University of Minnesota.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. http://dx.doi.org/10.2307/41410412, URL: https://www.jstor.org/stable/41410412 Publisher: Management Information Systems Research Center, University of Minnesota.

Villa, S., Kosch, T., Grelka, F., Schmidt, A., & Welsch, R. (2023). The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior*, *146*, Article 107787. http://dx.doi.org/10.1016/j.chb.2023.107787, URL: https://www.sciencedirect.com/science/article/pii/S0747563223001383.

Villa, S., Niess, J., Schmidt, A., & Welsch, R. (2023). Society's attitudes towards human augmentation and performance enhancement technologies (SHAPE) scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*(3), 1–23. http://dx.doi.org/10.1145/3610915, URL: https://research.aalto.fi/en/publications/societys-attitudes-towards-human-augmentation-and-performance-enh Publisher: ACM.

Wells, J. D., Campbell, D. E., Valacich, J. S., & Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: A risk/reward perspective. *Decision Sciences*, *41*(4), 813–843. http://dx.doi.org/10.1111/j.1540-5915.2010.00292.x, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2010.00292.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.2010.00292.x.

Welsch, R., Schmidt, A. F., Turner, D., & Rettenberger, M. (2021). Test–retest reliability and temporal agreement of direct and indirect sexual interest measures. *Sexual Abuse*, *33*(3), 339–360. http://dx.doi.org/10.1177/1079063220904354, arXiv:https://doi.org/10.1177/1079063220904354 Pmid: 32036763.

Wilson, M. L., Huggins-Manley, A. C., Ritzhaupt, A. D., & Ruggles, K. (2023). Development of the abbreviated technology anxiety scale (ATAS). *Behavior Research Methods*, *55*(1), 185–199. http://dx.doi.org/10.3758/s13428-022-01820-9.

Woźniak, P. W., Karolus, J., Lang, F., Eckerth, C., Schöning, J., Rogers, Y., & Niess, J. (2021). Creepy technology:What is it and how do you measure it? In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–13). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3411764.3445299, URL: https://dl.acm.org/doi/10.1145/3411764.3445299.

Wyatt, S. (2008). Technological determinism is dead; long live technological determinism. *The Handbook of Science and Technology Studies*, *3*, 165–180.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, *25*(5), 508–518. http://dx.doi.org/10.1177/0193945903252998, URL: https://journals-sagepub-com.emedien.ub.uni-muenchen.de/doi/10.1177/0193945903252998 Publisher: SAGE Publications Inc.