

A Scale to Assess Consumer Perceptions of Technology Product Inclusivity: Development and Validation

**Carmen Van Ommen,
M.S.**

Embry-Riddle Aeronautical
University
1 Aerospace Blvd.
Daytona Beach, FL, USA
carmen.vanommen@
my.erau.edu

**Barbara S. Chaparro,
Ph.D.**

Embry-Riddle Aeronautical
University
1 Aerospace Blvd.
Daytona Beach, FL, USA
barbara.chaparro@
erau.edu

**Joseph R. Keebler,
Ph.D.**

Embry-Riddle Aeronautical
University
1 Aerospace Blvd.
Daytona Beach, FL, USA
joseph.keebler@erau.edu

Sanjay Batra

Hardware Research
Manager
Google
1600 Amphitheatre Pkwy
Mountain View, CA, USA
batrasan@google.com

Mei Lu

UX Research Lead, Design
For All
Google
1600 Amphitheatre Pkwy
Mountain View, CA, USA
meilu@google.com

Abstract

Consumers are increasingly loyal to brands that are inclusive of them and their values, which promotes a sense of belonging. The goal of inclusive design is to design for the widest population possible; however, technology product inclusivity is a multifaceted concept that reflects consumer experience with a product, perceptions of trust, and ways to satisfy psychological needs. Technology products that foster a sense of belonging may promote perceptions of product quality, satisfaction, and usability. Although many measurement scales have been developed to assess consumer perceptions of technology products, no published scales exist to assess consumer perceptions of product inclusivity. The authors of this article adhered to best practices in scale development to generate an item pool based on a literature review, expert review, pilot studies, exploratory factor analysis (EFA; $N = 785$), and two confirmatory factor analyses (CFA; $N = 677$; CFA II; $N = 588$). We refined an initial pool of 194 items to a 25-item scale. Factor analyses indicate 5 factors contribute to the perceptions of product inclusivity: Personal Connection, Product Challenges, Confidence in Usage, Meets Expectations, and Company Empathy. The goal of developing this scale is to allow companies and product designers to measure how inclusive their technology product is, as well as to gain insight into areas of inclusivity that excel or could be improved.

Keywords

technology products, design, inclusivity, customer satisfaction, scale development, scale validation



Introduction

In recent decades, society has shifted toward inclusivity, that is, designing for a broader population, especially in consideration of older adults or people with disabilities (Clarkson & Coleman, 2010). This shift was largely furthered by Ron Mace, with his promotion of universal design that aimed to understand and respect the needs of a diverse range of users (Clarkson et al., 2013). Several other approaches, such as design for all or inclusive design, have also focused on achieving this target (Keates et al., 2000; Clarkson & Coleman, 2010).

When discussing what inclusive design means, previous publications have presented some of the following definitions:

- Inclusive design involves creating solutions that are accessible to and usable by as many persons as reasonably possible (Keates, 2005).
- Inclusive design ensures that every person—regardless of their gender, location, native language, and physical abilities—can enjoy and use products or services (Interaction Design Foundation, 2016).
- Inclusive design describes methodologies to create products that understand and enable people of all backgrounds and abilities (Kendrick, 2022).
- [Inclusive design involves] designing a diversity of ways for users to participate so that everyone has a sense of belonging (Holmes, 2018).

In general, the key aspect of inclusive design has been to design for the widest population possible while fostering a sense of enjoyment and belonging.

A researcher's ability to assess whether a user is included or excluded from using a product can be influential, especially in decision-making about product design within organizations. Although Keates and Clarkson (2002) proposed a method for evaluating design exclusion, there are fewer tools available for measuring feelings of inclusion. Currently, the methods for assessing inclusivity are time-consuming (self-observation, interviews, expert review, and user trials), yet quicker methods (questionnaires and checklists) have not been specifically designed for inclusivity or made suitable for a wide range of products (Clarkson et al., 2013). There is a need for flexible and quick assessments to be developed to allow designers and researchers to evaluate products quickly and frequently throughout the design process.

The use of technology in everyday life has become ubiquitous. Smartphones, for example, are the most frequently used consumer electronics (Bashir, 2024); they provide benefits to users, not only in terms of connecting one with others, but also servicing vulnerable populations by allowing older adults more independence, translating for people who do not speak the same language, and acting as assistive technology for people with disabilities. Because of the widespread usage and the possibility for technology to improve lives, it has become important to understand how users perceive the inclusivity of the technology so designers can ensure as much of the population as possible can use their products. A measurement tool for assessing technology product inclusivity can provide guidance on whether target user groups feel included as users of the product, to what degree they feel included, and which areas of the product design can be changed to improve feelings of inclusivity.

In this paper, the authors have reviewed constructs related to the inclusivity of technology products. Following the review, we have described our efforts to create and psychometrically validate a scale that measures technology product inclusivity, with the aim of using the scale to assist in the measurement of perceptions of technology product inclusivity (PTPI).

Measures Related to the Perception of Inclusivity

Aspects of product usage that contribute to consumers' sense of inclusivity have varied in their focus. For example, a primary factor may be the extent to which the product simply fulfils a consumer's psychological need or allows customization specific to their need. Another factor may include the consumers' feelings experienced while using the product, such as how enjoyable or engaging the product is, or how confident or in control the user feels. Additionally, it is possible that feelings of trust in the technology itself or the company that makes the product may contribute to feelings of inclusivity. The following sections summarize how these constructs have been measured and how they may apply specifically to technology products.

Personalization

Companies have increasingly discovered that consumers prefer products or services that allow them to express themselves, that are customized to their preferences or needs, and that are not standardized goods or services (Coelho & Henseler, 2012). Examples of customization have ranged from changing the color of the product to how the product operates. For example, a consumer may want a teal phone case, a PC they can customize with high-powered components for competitive gaming, or a website that recommends products based on previous purchases. Products or services that consumers customize can become more personal to them, thus creating a stronger emotional bond (Sashi, 2012). Several studies have shown that offering personalization for services has correlated with higher levels of customer loyalty and satisfaction (Coelho & Henseler, 2012; Ball et al., 2006; Bock et al., 2016). Additionally, offering a personalized experience in mobile shopping apps has been shown to increase the likelihood of impulse purchases (Chopdar et al., 2022).

Customers are more likely to purchase products that allow self-expression and help them align with their chosen social groups. Research has shown that Gen Z consumers often purchase products to show off "who they are" to others (Shin et al., 2021). Gen Z consumers may direct attention to the products they have purchased to facilitate a sense of belonging (Lee, 2020). Self-presentation or self-expression significantly influences the intent to repurchase products, such as those made by luxury brands (Kauppinen-Räsänen et al., 2018). Using product purchases as a form of self-expression has not been limited to Gen Z. Results from a survey by WP Engine™ showed that a range of generations, including 31% of Gen Z consumers, 38% of Millennials, 40% of Gen X, and 44% of Baby Boomers will stop purchasing products from brands that contribute to social causes that they do not believe in (Selig, 2020).

Customers want products that fit their individualized wants and needs and that allow for self-expression that fosters a sense of belonging within their preferred social groups. Otherwise, if there is a lack of customization options available, consumers may have to adapt to the design constraints of the product rather than the design of the product originating to meet their needs.

Psychological Needs

The Balanced Measure of Psychological Needs (BMPN) scale was developed to assess the satisfaction of psychological needs (Sheldon & Hilpert, 2012). The BMPN was based on Self-Determination Theory, which states that humans are motivated to grow by three basic needs: competence, relatedness, and autonomy (Ryan & Deci, 2000). Competence refers to the ability to effectively perform the task one does and to master skills. Relatedness refers to feeling close and connected to others, and autonomy refers to the ability to own one's experiences and to do things of one's own will. Research in this area has shown that needs fulfillment can be measured generally in a person's life experience and also in specific settings. Applying this to technology products, a consumer should be able to master their product use, feel close to others when using the product, and use the product of their own choosing, however they want.

Belonging

Malone, Pillow, and Osman (2012) developed a scale of general belongingness (GBS) that is measured by two related factors: Acceptance/Inclusion and Rejection/Exclusion. Both factors may represent a sense of belonging in a consumer. For example, an iPhone® smartphone user may report that their use of the phone creates a sense of belonging because it helps them to connect with others and facilitate bonds with family and friends. However, it may also make them feel as if they do not belong if they are in a group SMS chat with Android™ smartphone users and their interface and ability to interact, or react to posts, differ. It has become important that designers of technology products understand the design and interaction elements that contribute to perceptions of belonging.

Product Experience

In addition to needs fulfillment, researchers have developed scales to measure perceptions of product experience. These include aspects of the product design itself as well as perceptions of autonomy, enjoyment, and engagement with a product. Gilal et al. (2018) proposed the Four-Factor Model of Product Design, which measures user perceptions of a product's visual appeal

(Affective), quality and function (Cognitive), ease of use (Ergonomic), and personalization (Reflective).

Tapal et al.'s (2017) Sense of Agency scale was developed to measure consumers' sense of autonomy. A sense of autonomy is an important part of feeling that one is included or valued; therefore, autonomy is an important part of self-expression. If there is no sense of autonomy, a person cannot express themselves in a way that shows their true nature; instead, they express themselves in a way that they think aligns with others' expectations. Assessing a consumer's sense of autonomy during product experience has allowed us to measure whether the consumer feels valued, included, and able to express their true selves during that experience.

Davidson et al. (2022) measured enjoyment with the psychometrically validated ENJOY Scale. This scale was developed to measure enjoyment in an activity through five factors: Pleasure, Relatedness, Competence, Challenge/Improvement, and Engagement. Although this scale was not developed specifically for product evaluation, it has become evident that an enjoyable product experience can contribute to overall product satisfaction and perhaps feelings of inclusivity.

O'Brien and Toms (2009) measured self-reported engagement with the User Engagement Scale (UES), which was developed to measure six dimensions: Aesthetic Appeal, Focused Attention, Novelty, Perceived Usability, Felt Involvement, and Endurability. Providing a consumer with an engaging product experience goes beyond simple effectiveness or satisfaction; it provides an experience that draws in a consumer so they feel absorbed in a positive manner.

Trust

Researchers have measured trust in the context of technology (Chi et al., 2021), as well as in online technology companies (Bhattacharjee, 2002). Chi et al. (2021) determined that trust was an important factor in robot acceptance, in particular, trust within human-robot interactions and how well the robot met the users' needs and expectations. Bhattacharjee (2002) showed that consumers' trust in an online company predicted their willingness to engage with that company, with a contributing subfactor being their perceptions of company benevolence. If a consumer believes that the company that developed their technology product is well-meaning and caring, this could impact their feelings of inclusivity with the product.

Each of these constructs provides some level of understanding of a consumer's experience with a technology product, yet there are currently no validated scales that assess how the constructs collectively contribute to a consumer's perception of inclusivity when using the product. In addition, it is not known whether perceptions of inclusivity can be expressed similarly across a range of technology products (such as smartphones versus gaming devices or security systems).

Research Purpose

The purpose of this research was to develop a validated, comprehensive scale that reflects consumer perceptions of inclusivity for technology products. As the basis for scale development, we used research suggesting that customization, self-expression, psychological need, belonging, product experience, and trust influence technology product inclusivity and product satisfaction for a wide range of users who have varying backgrounds and capabilities.

Adhering to best practices for scale development, we conducted the following steps:

1. **Item Pool Generation:** We examined a broad selection of scales through an extensive literature review to determine particular items that could contribute to the measurement of inclusivity.
2. **Expert Review of Item Pool:** Experts in scale design and/or inclusivity, whom we consulted with, rated the importance and relevance of each potential scale item.
3. **Questionnaire Pilot Study:** We piloted the scale with participants of varied demographics and capabilities.
4. **Exploratory Factor Analysis (EFA):** We distributed a refined scale to participants in an online survey ($N = 785$) and conducted an EFA to identify underlying factors.
5. **Confirmatory Factor Analyses (CFA):** We surveyed an independent sample of participants ($N = 677$) using the revised scale from the EFA to further explore and

validate the constructs. Finally, we conducted a second CFA ($N = 588$) to further validate the final 25-item scale.

The goal of developing this scale was to allow companies and product designers to measure how inclusive their technology product is, as well as to gain insight into areas of inclusivity that excel or could be improved.

Scale Development

Step 1: Item Pool Generation

We conducted a literature review to identify scholarly research and published scales that measured constructs relating to consumer perceptions of products with a particular interest in technology products. We searched Google™ Scholar™ using keywords such as “accessibility,” “disability,” “product,” “technology,” “evaluation,” “scale,” “inclusion,” “inclusivity,” “belonging,” “diversity,” and “trust,” and we found 430 related references. After reviewing the abstracts, we selected 59 as appropriate to include. These references included research related to user perceptions regarding product design, engagement, perceived usability, and such. In addition, we examined research articles, books, and published scales that reflected perceptions of inclusivity and belonging.

We selected 11 scales (Table 1) that measured aspects we considered particularly related to technology product inclusivity. The combined scale items from these questionnaires formed the initial draft of the scale. Additionally, 16 scale items were added that reflected findings from published research related to inclusivity. In total, 204 scale items were generated from these sources.

We assessed the collection of items to reduce redundancy. Items that were too similarly phrased were removed (for example, “I lost track of what was going on outside of using the product” and “I lost track of what was going on around me while using the product”). Additionally, items that were related to a consumer’s general frame of mind that were not applicable to a product were removed (for example, “I join groups more for friendship than the activity itself”). Scale items were also examined by a technical writer using the Flesch-Kincaid reading ease formula (Kincaid et al., 1975) to ensure the wording was appropriate and easy to understand. All scale items were modified to reference technology products. For example, the statement “I have a sense of belonging” was modified to “I have a sense of belonging when using this product.” We examined the scale items to ensure they were appropriate for a wide range of technology products, rather than a specific product or product category (such as Wi-Fi routers, thermostats, smart devices, gaming devices, software, or vehicle technology, etc.). After modifying and refining the potential scale items, 125 items were retained for further consideration in an expert review. Table 1 shows each scale and the number of items taken from each source.

Table 1. Overview of the Number of Items Derived From Each Source

Source	Name of Questionnaire	No. of Items
Sheldon and Hilpert (2012)	Balanced Measure of Psychological Needs Scale (BMPN)	13
Malone et al. (2012)	General Belongingness Scale	7
Gilal et al. (2018)	4-Factor Model of Product Design	14
Tapal et al. (2017)	Sense of Agency Scale (SoAS)	7
Davidson et al. (2022)	ENJOY Scale	20
O'Brien and Toms (2009)	User Engagement Scale	14
Chi et al. (2021)	Trust with AI Social Service Robots	8

Source	Name of Questionnaire	No. of Items
Bhattacharjee (2002)	Trust in Online Technology Companies	3
O'Brien et al. (2018)	User Engagement Scale - Short Form (UES-SF)	11
Lee and Robbins (1995)	Measuring Belongingness: Social Connectedness and Social Assurance Scales	9
Romansky et al. (2021)	Workplace Inclusion	3
Scholarly research literature	N/A	16

Steps 2 and 3: Expert Review of the Item Pool and a Questionnaire Pilot Study

Expert Review

A critical part of scale development is ensuring content validity, that is, that the scale is measuring the construct desired. A technique to help improve content validity is to have an expert review. Experts in the construct domain review scale items and give feedback about whether the items are applicable to the construct being measured, whether the items are understandable, if the items need to be revised, or if new items need to be added (Worthington & Whittaker, 2006).

In our research, 12 participants reviewed the item pool. Participants were recruited by word of mouth and were required to be experts in psychometric questionnaire development or inclusive design to receive an invitation to participate. Overall, 5 participants were experts in questionnaire development, 5 participants were experts in inclusive design, and 2 participants were experts in both areas.

The purpose of the expert review was to gather feedback to inform and improve the design of the new inclusivity scale. We administered the questionnaire using Qualtrics™ Online Survey Software to experts to give feedback on scale items and capture comments. The questionnaire contained a series of 125 statements from the generated item pool. The participants were asked to select a technology product they used frequently and provide information on their familiarity with, frequency of use, and satisfaction with the product. Examples of technology products were provided, including IoT devices (such as app-connected locks or doorbells), computers, self-driving cars and vehicle technology, televisions and smart screens, smart speakers, wearables (such as smart watches or 911-alerting products), gaming devices, drones and robotics, audio and music equipment, communication devices, mobile devices, and software. Participants then progressed to the evaluation, where they were asked to rate their experience with the technology product using a seven-point Likert Scale (1 = Strongly disagree, 7 = Strongly agree). Participants also rated whether each item was relevant to the topic of inclusivity (Yes, No, or Unsure) and how important they thought it was to include the item in a measure of product inclusivity (High, Medium, or Low). After reviewing a set of 5 items, participants gave their comments or suggestions for rewording any of the items. After all 125 items were reviewed, participants were able to provide additional items or content areas that they felt were important to measuring product inclusivity, which were not included in the questionnaire. The expert review was estimated to take 2 hr to complete, and participants were offered \$100 or company perks upon completion of the questionnaire.

After an analysis of expert feedback, the total item pool was reduced to 79. Fifty-one items were removed from the pool, 5 items were added, and the wording of 13 items was modified for clarity. The remaining items were used in the pilot study.

PILOT STUDY

We conducted a pilot study with 10 participants. Sessions were conducted face-to-face and remotely over video conference (such as Zoom™) or over messaging software (such as Discord™) for the participant who was hard of hearing. Participants included those who self-identified as having a disability, those who were over 60 years old, and those whose native language was not English. Of the participants, half ($n = 5$) were females, 4 were non-native English speakers, 1 was over the age of 60, and 4 identified as having a disability. Of the non-native English speakers, all were fluent in English. Of the participants who identified as having a disability, 1 had an auditory disability, 1 had a visual disability, 1 had a physical disability, and 1 had a cognitive/mental disability. This group of pilot participants was purposefully selected to ensure that all items on the questionnaire could be easily understood and were applicable to a diverse audience. Pilot participants were recruited by word of mouth and had not previously completed the survey. During the pilot study, participants were asked about the clarity of the statements and to share their interpretation of the statements.

MATERIALS

Qualtrics, an online survey tool, was used to create the questionnaire and capture responses. The questionnaire contained the series of statements retained and modified after the expert review, which were rated on a seven-point bipolar scale with response anchors (Vagias, 2006). The response options included a not applicable (N/A) option at the end.

PROCEDURE

Before evaluating the technology product, participants were asked to enter its make and model. Participants also provided usage information about the technology product, including how long they had owned or used the product, how often they used it, how familiar they were with it, and their general satisfaction. Participants then proceeded to the evaluation phase and indicated, on a seven-point rating scale, their level of agreement with each statement about their experience with the technology product.

We conducted the pilot study both in-person and remotely using video conference or messaging software. Participants were asked to think aloud when reading the statements and indicate when they encountered words or statements that were difficult to interpret or that were not fully representative of their experience. After finishing the questionnaire, participants were asked to provide final comments about the questionnaire, including whether the questionnaire was able to fully capture their experience in terms of their sense of inclusivity in relation to the technology products they used.

In addition, the statements were reviewed by a technical writer.

RESULTS

Analysis of the pilot study data resulted in the revision of 58 items. Most revisions were made to change the tense of the items (from past to present tense), and the wording of some items was simplified or clarified. No items were added or removed after the pilot study, keeping the total number of items at 79. Comments collected from participants and the technical writer during this phase were used to improve the comprehension of the statements.

Steps 4 and 5: Exploratory Factor Analysis, Confirmatory Factor Analysis I, and Confirmatory Factor Analysis II

Methodology

A total of 1160 surveys were collected in the EFA study, 767 were collected in the CFA I, and 718 were collected in the CFA II study.

During the data screening and cleaning process, 31.5% ($n = 375$) of surveys in the EFA study, 11.7% ($n = 90$) of surveys in the CFA I study, and 18.1% ($n = 130$) of surveys in the CFA II study were identified as containing nonvalid responses. Nonvalid responses consisted of participants who did not follow survey instructions (such as not evaluating a technology product, evaluating more than one technology product, or submitting multiple responses from the same IP address). Survey responses were also removed if they met one or more of the following pre-established criteria: a) finished the survey in less than 5 minutes (for the EFA

study), b) answered validation questions incorrectly, or c) answered with the same or similar responses (such as neither agree nor disagree throughout the whole survey).

PARTICIPANTS

Participants were recruited using Amazon™ Mechanical Turk and Prolific™. Participants were required to be at least 18 years of age and live in the United States. Participants were balanced by gender, and a representative sample of older adults, people with disabilities, and different ethnicities was recruited. Table 2 provides a summary of participant demographics for both studies. In total, 785, 677, and 588 valid responses were retained for the EFA, CFA I, and CFA II, respectively. Survey respondents for both studies had similar demographics. The EFA included 16.92% older adults, 12.72% participants reporting a disability, and 50% people of color. The CFA I had 16.69% older adults, 12.56% participants reporting a disability, and 52.88% people of color. The CFA II had 15.82% older adults, 13.44% participants reporting a disability, and 56.97% people of color. On average, participants were satisfied with their technology product ($M_{EFA} = 6.01$, $SD_{EFA} = 1.037$; $M_{CFA I} = 6.04$, $SD_{CFA I} = 1.054$; $M_{CFA II} = 6.02$, $SD_{CFA II} = 1.067$) and felt that they were included as a user of the product ($M_{EFA} = 5.59$, $SD_{EFA} = 1.152$; $M_{CFA I} = 5.58$, $SD_{CFA I} = 1.194$; $M_{CFA II} = 5.57$, $SD_{CFA II} = 1.108$).

Table 2. Demographics of Participants in the EFA ($N = 785$), CFA I ($N = 677$), and CFA II ($N = 588$) Studies

Variable	EFA Value	CFA I Value	CFA II Value
Age in Years ($M \mp SD$)	39.84 \mp 24.22	39.45 \mp 15.30	37.76 \mp 15.63
Age range in years	18-78	18-79	18-78
Gender (%)			
Male	50.13%	50.81%	51.02%
Female	46.82%	47.27%	46.60%
Transgender	0.38%	0.30%	0.34%
Nonbinary	1.27%	1.48%	2.04%
Prefer to self-describe	0.25%	0.00%	0.00%
I prefer not to answer	0.89%	0.15%	0.00%
Ethnicity (%)			
American Indian/Alaskan Native	2.29%	1.03%	0.85%
Asian/Pacific Islander	16.79%	19.79%	20.24%
Black/African American	14.89%	17.28%	18.03%
Hispanic/Latino	8.40%	7.39%	10.71%
White (not of Hispanic origin)	46.69%	46.23%	42.35%
Biracial/multiracial/mixed	7.63%	7.39%	7.14%
I prefer not to answer	1.91%	0.30%	0.68%
Education Level (%)			
Highschool	9.54%	13.41%	14.80%

Variable	EFA Value	CFA I Value	CFA II Value
Some college	18.58%	21.27%	18.54%
Associates degree	11.70%	11.08%	8.33%
Vocational/Technical college	3.44%	2.66%	1.70%
Bachelor's degree	40.97%	37.08%	43.03%
Master's degree	12.21%	11.96%	11.56%
Doctorate or professional degree	2.54%	3.25%	1.70%
I prefer not to answer	0.89%	0.30%	0.34%
Self-Identify as Having a Disability			
Yes*	12.72%	12.56%	13.44%
Physical	6.63%	7.53%	5.95%
Visual	1.45%	2.07%	2.21%
Auditory	1.78%	1.18%	1.36%
Cognitive/Mental	5.35%	5.32%	6.8%
Emotional	4.59%	3.4%	3.06%
Other	0.64%	0.3%	0.34%
No	83.33%	85.67%	83.33%
I prefer not to answer	3.94%	1.62%	3.23%

Participants could select more than one disability type, so percentages of disability types may sum to more than the total percentage of disability.

TECHNOLOGY PRODUCTS

Most of the technology products participants chose to evaluate were used daily and had been purchased 6-12 months prior to taking the survey. Given that the survey instructions asked participants to choose a technology product that they use frequently, it was expected that many of these products would be products that participants liked. The mean ratings for overall satisfaction (1 = Extremely dissatisfied, 7 = Extremely satisfied), confirmed these expectations, as most respondents chose to evaluate products that they liked. Table 3 provides a summary of the technology products evaluated in the EFA and CFA studies. Overall, the technology products evaluated in both studies covered 27 categories containing a variety of products within those categories.

Table 3. Overview of the Technology Product Categories Represented in the EFA ($N = 785$), CFA I ($N = 677$), and CFA II Studies ($N = 588$)

Technology Product Category	EFA (%)	CFA I (%)	CFA II (%)
Smartphone	28.03%	25.11%	26.33%
Computer	15.92%	14.48%	15.50%
Smart speaker	9.94%	9.16%	9.17%
Gaming device	9.04%	9.31%	9.51%
Smart watch	8.28%	7.98%	7.95%
Tablet	3.95%	4.28%	4.10%
Smart TV	3.69%	6.35%	5.12%
Other	3.57%	3.10%	3.27%
Headphones	3.31%	3.69%	3.75%
Smart doorbell	2.68%	1.62%	2.10%
Home products	1.91%	2.95%	2.24%
Software	1.66%	1.48%	1.71%
Speaker/Sound system	1.27%	0.59%	0.98%
Computer accessories	1.15%	1.62%	1.32%
Drone	0.89%	0.15%	0.44%
Smart TV remote	0.89%	0.59%	0.59%
TV accessories	0.89%	0.00%	0.39%
Smart display	0.64%	0.15%	0.39%
Smart thermostat	0.64%	0.00%	0.39%
Vehicle	0.64%	0.44%	0.59%
Security system	0.38%	1.77%	1.22%
Drawing tablet	0.25%	0.30%	0.20%
eReader	0.25%	1.18%	0.59%
Tool	0.13%	0.44%	0.20%
Audio equipment	0.00%	0.15%	0.34%
Exercise equipment	0.00%	0.30%	0.10%
Fitness tracker	0.00%	2.66%	1.56%

MATERIALS

Qualtrics Online Survey Software outputted the questionnaire for both the EFA and CFA studies. The questionnaire included the following sections:

1. Consent form
2. Make and model of the technology product under evaluation (participants entered the name in a text field)
3. Basic questions about familiarity and frequency of use of the product
4. Product evaluation statements
 - a. These statements were randomized, and five statements per screen were displayed to minimize scrolling.
 - b. Each statement was evaluated on a seven-point rating scale (1 = Strongly disagree, 7 = Strongly agree; N/A option at the end of the scale).
5. Overall satisfaction rating (1 = Extremely dissatisfied, 7 = Extremely satisfied)
6. Overall inclusion rating ("Overall, I feel included as a user of this product."; 1 = Strongly disagree, 7 = Strongly agree)
7. Likelihood-to-Recommend (LTR) (only included in the CFA studies)
8. System Usability Score (SUS; Brooke, 1996) (only included in the CFA studies)
9. Basic demographic questions (such as age, gender, and disability)

PROCEDURE

We shared information about the EFA study, including the survey link, on LinkedIn™, email lists, and various popular social networking websites (such as Reddit™). The EFA, CFA I, and CFA II surveys were also distributed on the survey distribution platforms Amazon Mechanical Turk and Prolific. The survey links were open for 60 days for the EFA study, 7 days for the CFA I study, and 13 days for the CFA II study. All participants who completed the EFA survey outside of the survey distribution platforms and who left their contact information were eligible to be entered into a raffle to receive one of 10 \$50 Amazon™ gift cards. Participants were informed that their email addresses would be used only for the purpose of selecting gift card winners.

EFA Results

NORMALITY

Inspection of the histograms and results of the Shapiro-Wilk test revealed that the majority of the items deviated significantly from a normal distribution. Most items were negatively skewed due to participants tending to give positive ratings about the technology product they chose (skewness $<|2|$ and kurtosis <7 ; Finney & DiStefano, 2013). This is consistent with most participants reporting that they were satisfied with the technology product. Due to the exploratory nature of this analysis, the data was not transformed.

MISSING DATA

N/A responses were treated as missing data. In total, 1.62% of the data was missing. Less than 5% of missing data is considered to be a small amount (Bennett, 2001; Schafer, 1999; Tabachnick & Fidel, 2013). Little's MCAR test results ($\chi^2 = 16250.936$, $df = 14392$, $p < .01$) indicated that the data was not missing completely at random. Because participants were allowed to indicate that an item was not applicable to their experience, and N/A responses were treated as missing data, it is possible that the data is not missing completely at random. Certain items were more likely to have a response of not applicable; 76.2% of variables or scale items ($n = 76$) and 33% of cases or participants ($n = 263$) contained at least one missing value. The percentage of missing values for each variable ranged from 0.13% to 11.07%. Because all of the variables contained less than 20% of missing values, none were removed from the initial stage of data analysis.

Because the missing data in this data set was not missing completely at random, we applied the expectation maximization method (EM). This method models or estimates the parameters of the data, then replaces the data to fit those parameters. Roth (1994) states that replacing missing values based on models or parameter estimates was more accurate for data that was

systematically missing. Additionally, if more than 10% of the data is missing, selecting an appropriate method to replace missing values becomes more important (Tsiriktsis, 2005).

FACTORABILITY

To determine whether a factor analysis was appropriate to analyze the data, multiple criteria were used, including adequate sample size, correlation matrix, Kaiser-Meyer-Olkin (KMO), Bartlett's test of sphericity, anti-imaging correlation matrix, communalities, and factor loadings. When determining the adequacy of a sample size, researchers generally agree that a sample size of at least 300 is desirable to perform a factor analysis (Worthington & Whittaker, 2006; Tabachnick & Fidell, 2013; MacCallum et al., 1999). Comrey and Lee (1992) provide classifications of sample sizes, stating that a size of 100 is "poor," 300 is "good," and 500 is "very good." Thus, the sample size for this study ($n = 785$) was deemed suitable and in the "very good" range.

We also evaluated the correlation matrix between items to determine if the use of factor analysis was appropriate. According to Tabachnick and Fidell (2013), the majority of items should have intercorrelations above $|.30|$ and below $|.90|$. Results showed that the majority of items in this data set had correlations above $|.30|$ and no items had correlations above $|.90|$, so the use of factor analysis was still appropriate at this point. No items were removed during this phase of the analysis.

The Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity results also suggest that a factor analysis is appropriate. The KMO indicates whether results should generate distinct and reliable factors, with .6 or above being the minimum required for an analysis (Worthington & Whittaker, 2006). The KMO for this data set was .97. Significant results from Bartlett's test of sphericity suggest that intercorrelations are due to common variance between items (Worthington & Whittaker, 2006). Results from this data set were significant ($\chi^2 = 13974.420$, $df = 465$, $p < .005$), which further supports factor analysis.

Finally, the anti-image correlation matrix, communalities, and factor loadings were examined to evaluate scale factorability. When evaluating the anti-image correlation matrix, all items were .8 or above, which is well above the suggested cutoff point of .50 (Worthington & Whittaker, 2006), thus, none of the items were removed at this point. Results also showed that the majority of the items had communalities above .50, and each factor contained at least 3 items with factor loadings above $|.50|$. After evaluating all indicators as well as taking into account a sample size of over 750, the results contribute to an overall confidence that conducting a factor analysis is appropriate (Worthington & Whittaker, 2006).

FACTOR EXTRACTION, ROTATION, AND RETENTION

Because the data was not normally distributed, we chose principal axis factoring as the extraction method (Costello & Osborne, 2005). In terms of the rotation method, Costello & Osborne (2005) suggested that orthogonal rotations, which assume that factors are not correlated, should be used because it is easier to interpret the results. However, factors are likely to be correlated in social sciences, so using an oblique rotation can reveal valuable information about these correlations and produce a more accurate and reproducible solution. Using this argument, an oblique rotation was chosen. Based on recommendations from the literature (Fabrigar et al., 1999; Matsunaga, 2010; Russell, 2002), we selected the promax rotation ($kappa = 4$).

Multiple factor retention strategies helped determine the number of factors to retain. Kaiser (1958) suggested that factors with eigenvalues less than 1 are potentially unstable and thus should not be retained. Based on this criterion, there were 11 factors to retain. An additional strategy included Cattell's scree test (Cattell, 1966). Visual inspection of the scree plot suggested a three-factor solution. Horn's parallel analysis was also used to determine the number of factors to retain. Results obtained from the parallel analysis revealed that there were 6 underlying factors.

In addition to the three strategies mentioned previously, other criteria were used to inform the process of factor retention. Factors with fewer than 3 items were considered unstable and would therefore be rejected (Costello & Osborne, 2005; Hinkin, 1995). Additionally, factors should typically have a simple structure and be easy to explain, so factors that were difficult to interpret or understand would not be retained.

When interpreting the factors, we examined both the pattern matrix and structure matrix. The primary focus was on the pattern matrix because it displays clearer results about which items load uniquely onto which factor (Tabachnick & Fidell, 2013; Costello & Osborne, 2005). The $|0.40|$ value was the cutoff value for item loading because it is the most common value and it is in the recommended range of cutoff values (Tabachnick & Fidell, 2013; Hinkin, 1995; Nunally, 1978).

ITEM REMOVAL

Item removal improved the interpretability of the data and factor structure. Using multiple criteria, we determined to remove items that contained factor loadings below $|.32|$, cross-loaded on 2 or more factors with a difference of greater than $|.15|$, made little or no contribution to the internal consistency of the scale, had low conceptual relevance to a factor, or were not conceptually consistent with other items loaded on the same factor (Worthington & Whittaker, 2006). An internal reliability analysis (Cronbach's alpha) and EFA were run each time an item was deleted to ensure that the item removal would not have a major effect on the factor structure or the internal consistency of the scale.

After applying this criterion, 48 items were removed from further analysis. Cronbach's alpha for the remaining 31 items was .915. This exceeded the acceptable threshold of .70 (Hinkin, 1995; Nunally, 1978).

FACTOR SOLUTION

Following item removal, the five-factor solution was determined to be the simplest and most conceptually relevant solution. The five-factor solution explained approximately 63.18% of the total variance, and Cronbach's alpha for each factor or subscale exceeded the .70 acceptable threshold (Hinkin, 1995; Nunally, 1978). Table 4 summarizes the eigenvalues and Cronbach's alpha results for each factor. We named the 5 factors Confidence, Personal Connection, Meets Expectations, Product Challenges, and Company Empathy. To better understand the relationship between factors, we conducted Pearson's correlation tests. Results indicated that there is a significant positive relationship between all of the factors except for Personal Connection and Product Challenges. Participants indicated during the questionnaire their feeling of inclusivity, or whether they felt included as a member of the target audience for that product. The results of that item were also correlated with the factors. Table 5 displays the correlation results between factors and their relationship with a feeling of inclusivity.

Table 4. Five-Factor Solution: Summary of Eigenvalues and Cronbach's Alpha

Factor Number	No. of Items	Eigenvalues	Percentage of Variance Accounted for per Extracted Factor	Cronbach's α
Factor 1: Personal Connection	8	10.007	32.281	0.904
Factor 2: Product Challenges	9	5.530	17.840	0.897
Factor 3: Confidence in Usage	5	1.790	5.775	0.878
Factor 4: Meets Expectations	6	1.387	4.473	0.894
Factor 5: Company Empathy	3	.871	2.811	0.79

Table 5. Correlations Between Factors and With Inclusivity

	Factor 1: PC	Factor 2: CH	Factor 3: C	Factor 4: ME	Factor 5: CE	Overall Perception of Inclusivity
Factor 1: PC	1					
Factor 2: CH	-0.061	1				
Factor 3: C	.338**	.451**	1			
Factor 4: ME	.520**	.412**	.644**	1		
Factor 5: CE	.602**	.150**	.347**	.627**	1	
Overall Perception of Inclusivity	.612**	.136**	.409**	.581**	.623**	1

PC = Personal Connection; CH = Product Challenges; C = Confidence in Usage; ME = Meets Expectations; CE = Company Empathy. ** $p < .01$ (two-tailed).

CFA Results

During this step of the scale validation, two CFA studies were conducted. The first CFA (CFA I) was conducted to confirm hypothesized model results from the EFA study and to compare model fit with other proposed models. A model was chosen based on these CFA results. After reviewing the items in the model, 1 item was removed, and slight wording changes were made to 3 other items, so a second CFA (CFA II) study was conducted to ensure that there were no negative impacts on model fit.

NORMALITY

Similar to the EFA study, the histograms and results of the Shapiro-Wilk test for both the CFA I and CFA II revealed that the majority of the items deviated significantly from a normal distribution. Most items were considered moderately negatively skewed (skewness $< |2|$ and kurtosis < 7 ; Finney & DiStefano, 2013). Again, the decision was to not transform the data but keep it in its original state.

MISSING DATA

N/A responses were treated as missing data during both studies. In total during the CFA I, 1.15% of the data was missing. Little's MCAR test results ($\chi^2 = 2521.036$, $df = 2156$, $p = .000$) indicated that the data was not missing completely at random; 96.8% of variables or scale items ($n = 30$) and 18.6% of cases or participants ($n = 126$) contained at least one missing value. The percentage of missing values for each variable ranged from 0.15% to 8.86%.

During the CFA II, 1.02% of the data was missing. Little's MCAR test results ($\chi^2 = 1946.352$, $df = 1381$, $p = .000$) indicated that the data was not missing completely at random; 84% of variables or scale items ($n = 21$) and 13.6% of cases or participants ($n = 80$) contained at least one missing value. The percentage of missing values for each variable ranged from 0.17% to 4.08%.

The expectation maximization method was once again used to replace missing data in both the CFA and CFA II. Because there was less than 10% of data missing overall and for each variable, the difference in the end result between methods for replacing data is not as great; however, expectation maximization was used as it is more appropriate for use when data is not missing completely at random (Roth, 1994).

MODEL FIT ASSESSMENT

To evaluate model fit, it is recommended to use two to three fit indices, alongside the chi-square test statistic (Worthington & Whittaker, 2006; Hu & Bentler, 1999). The three fit indices we used were the Comparative Fit Index (CFI; Bentler, 1990), root mean square error of approximation (RMSEA; Steiger, 1980), and the Tucker Lewis Index (TLI; Tucker & Lewis,

1973). This is because chi-square has been widely criticized based on its assumption that the model fits perfectly with the population and its sensitivity to larger sample sizes and non-normal data. Thus, reporting the chi-squared test statistic is suggested, but it should not be depended on for assessment of overall model fit.

HYPOTHESIZED FACTOR MODEL FIT ASSESSMENT

The five-factor solution resulting from the EFA study was used in the CFA I study as the hypothesized full model. The model consists of the following unobserved latent factors: Confidence in Usage (5 items), Personal Connection (8 items), Meets Expectations (6 items), Product Challenges (9 items), and Company Empathy (3 items). Each of the items is considered an observed or measured variable in CFA.

Results showed that the hypothesized five-factor model has an overall good fit with the new data sample. The chi-square statistic ($\chi^2 = 2861.112$, $df = 848$, $p = .000$) was significant, which is likely due to the large sample size. The three primary goodness-of-fit statistics (CFI, RMSEA, and TLI) suggest that there is a good fit between the full five-factor model and the observed data (Table 6).

Table 6. Hypothesized Five-Factor Model's Fit Statistics ($N = 677$)

Fit Index	Value	Recommended Cutoff Values for Acceptable Fit	References
χ^2	$\chi^2 = 2861.112$, $df = 848$, $p < .005$	N/A	N/A
CFI	.927	0.90-0.95, > 0.95 indicates good fit	Hu and Bentler (1999) Bentler (1990)
RMSEA	0.039 [0.037, 0.040]	0.06-0.08, < .06 indicates excellent fit	Costello and Osborne (2005) Tucker and Lewis (1973)
TLI	.919	0.90-0.95, > 0.95 indicates good fit	Hu and Bentler (1999)

MODEL COMPARISON

During the CFA I, the hypothesized five-factor model was compared against two alternative models in terms of overall model fit. All models had the same number of cases ($N = 677$). The first alternative model consisted of 4 factors, based on eigenvalues greater than 1 as suggested by the EFA results, which removed the Company Empathy factor. The second alternative model utilized a five-factor model but had fewer items ($N = 27$), due to items being removed due to redundancy and conceptual relevance, and allowed items to covary.

All models had acceptable CFI, RMSEA, and TLI values and yielded statistically significant results in the chi-square difference test. Overall, the results from the goodness-of-fit statistics suggest that all models could be appropriate. The five-factor model was the most appropriate model due to being the most conceptually appropriate and having fewer items, which allows for quicker testing when administering the scale. Table 7 presents the main fit statistics for all three models.

Table 7. Comparison of Chi-Square, AIC, and CFI Fit Indices Across Models ($N = 677$)

Model	χ^2	AIC	CFI
5 factors (uncorrelated)	$\chi^2 = 2861.112, df = 848, p < .005$	3149.112	.927
5 factors (correlated, items removed)	$\chi^2 = 957.262, df = 310, p < .005$	1147.262	.941
4 factors (removed Company Empathy)	$\chi^2 = 1144.307, df = 318, p < .005$	1264.307	.926

CFA II

After conducting the CFA I, minor wording changes were made to enhance clarity, and 2 items were removed to reduce redundancy ("This product represents my sense of style" in Factor 1, and "For this product to work, I had to make changes to my environment beyond my expectations" in Factor 2). A second CFA ($n = 588$) was conducted to ensure the wording changes and item removal did not negatively impact the strength of the model. With the exception of the items removed, the model was created with the same correlations as the model deemed most appropriate after the CFA I. Results from the CFA II were similar to the results from the CFA I, with all model fit indices above the acceptable threshold. The model created and validated after the CFA II is the final model (Figure 1).

Table 8. Comparison of Chi-Square, AIC, and CFI Fit Indices Across CFA I ($N = 677$) and CFA II Models ($N = 588$)

Model	χ^2	AIC	CFI	RMSEA	TLI
CFA I	$\chi^2 = 957.262, df = 310, p < .005$	1147.262	.941	.056 [.052, .060]	.933
CFA II	$\chi^2 = 812.734, df = 262, p < .005$	988.734	.940	.060 [.055, .065]	.931

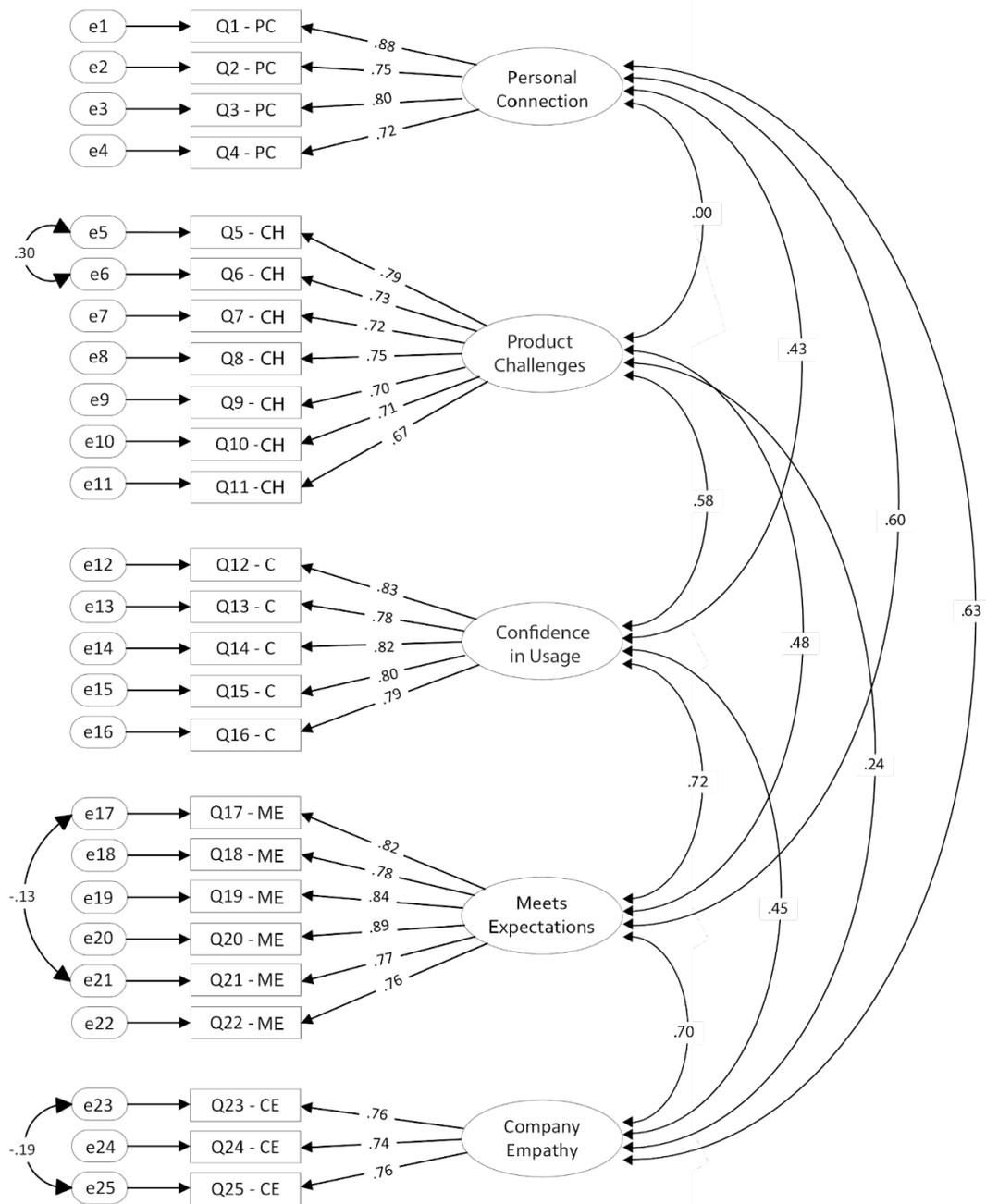


Figure 1. The Final Technology Product PTPI Scale model.

Scale Reliability and Validity Assessment

After the assessment of model fit, we assessed the reliability as well as the convergent and discriminant validity of the scale. The AMOS plugin, Master Validity Measures, was used to assess reliability and validity (Gaskin et al., 2019). The results revealed no validity concerns. Reliability values for all factors or subscales were above the suggested threshold of .7. Average variance extracted (AVE) values, or values representing convergent validity, were above the suggested threshold of .5 (Hair et al., 2010). All maximum shared variance (MSV) values, or

values representing discriminant validity, were less than AVE values. Thus, although there is variance shared between factors, the variance shared between items in a factor is less than the variance shared between factors. This suggests that the scale as a whole is measuring a construct, and each factor is measuring different dimensions of that construct. In sum, the results in Table 9 demonstrate that the final five-factor model has adequate reliability as well as convergent and discriminant validity.

Table 9. Reliability and Convergent and Discriminant Validity Results

	CR	AVE	MSV
Factor 1: Personal Connection	0.868	0.623	0.398
Factor 2: Product Challenges	0.886	0.527	0.339
Factor 3: Confidence in Usage	0.901	0.646	0.526
Factor 4: Meets Expectations	0.92	0.659	0.526
Factor 5: Company Empathy	0.794	0.562	0.493

CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance

Relationship to SUS and LTR

The final 5 factors were also correlated using Pearson's correlation test with the System Usability Scale (SUS) and Likelihood-to-Recommend (LTR) (Table 10) to better understand the relationship between the variables. Results indicated a positive relationship between all of the factors and the SUS and LTR. Meets Expectations showed a strong correlation with the LTR. Product Challenges and Confidence in Usage were strongly correlated with the SUS. Weaker relationships were found between Personal Connection and Company Empathy and the SUS, highlighting the unique insights the scale provides to product evaluation.

Table 10. Factor Correlations With SUS and LTR

	Factor 1: PC	Factor 2: CH	Factor 3: C	Factor 4: ME	Factor 5: CE	SUS	LTR
Factor 1: PC	1						
Factor 2: CH	-0.003	1					
Factor 3: C	.380**	.518**	1				
Factor 4: ME	.550**	.438**	.665**	1			
Factor 5: CE	.539**	.188**	.369**	.620**	1		
SUS	.278**	.752**	.726**	.701**	.416**	1	
LTR	.536**	0.300**	.414**	.743**	.596**	.504**	1

PC = Personal Connection; CH = Product Challenges; C = Confidence in Usage; ME = Meets Expectations; CE = Company Empathy. ** $p < .01$ (one-tailed).

To further explore the relationship between the 5 factors and the SUS and LTR, a structural equation model was created. The results of the model are displayed in Figure 2. The structural model revealed that the inclusivity factors significantly predicted SUS with a squared multiple correlation of .709, indicating that approximately 71% of the variance in perceived system

usability was explained by the inclusivity factors included in the model. According to Cohen's (1988) guidelines, this represents a substantial level of explained variance.

Inclusivity factors also significantly predicted LTR, with a .62 squared multiple correlation, suggesting that 62% of the variance in LTR was explained by the model's predictors. This R^2 value again indicates a substantial level of explained variance.

The subscale of Product Challenges was the highest contributor to the SUS score ($\beta = .66, p < .001$), whereas the subscale of Meets Expectations was the highest contributor to LTR ($\beta = .70, p < .001$).

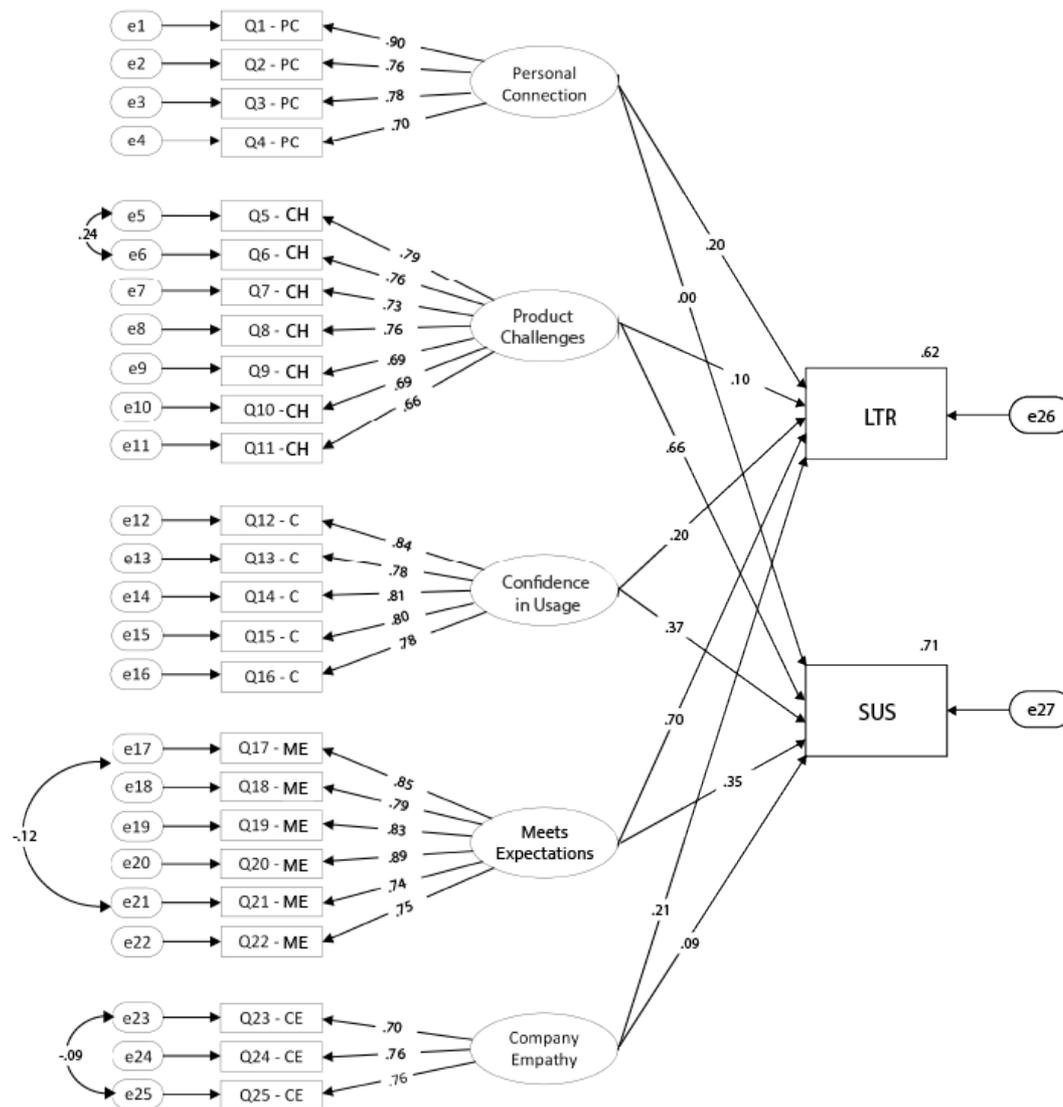


Figure 2. SEM exploring the relationship of subscales to SUS and LTR.

Model fit results were obtained for the chi-square statistic, CFI, RMSEA, and TLI (Table 11). The chi-square statistic ($\chi^2 = 2100.416, df = 313, p < .005$) was significant, which is likely due to the large sample size, and the fit indices suggested a less optimal fit.

Table 11. SEM Including LTR and SUS Model Fit Statistics ($N = 588$)

Fit Index	Value	Recommended Cutoff Values for Acceptable Fit	References
χ^2	$\chi^2 = 2100.416, df = 313, p < .005$	N/A	N/A
CFI	.834	0.90-0.95, > 0.95 indicates a good fit	Hu and Bentler (1999) Bentler (1990)
RMSEA	0.099 [0.095, 0.103]	0.06-0.08, < .06 indicates excellent fit	Costello and Osborne (2005) Tucker and Lewis (1973)
TLI	.814	0.90-0.95, > 0.95 indicates a good fit	Hu and Bentler (1999)

Conclusion

This study aimed to create and validate a measure of technology product inclusivity. Results from a rigorous process of scale development and validation revealed a new scale consisting of 25 items with 5 subscales: Confidence in Usage, Personal Connection, Meets Expectations, Product Challenges, and Company Empathy. Table 12 provides a brief description of each subscale. The scale is called the Perceptions of Technology Product Inclusivity (PTPI).

Table 12. Description of Each Subscale

Subscale	Description
Confidence in Usage	Perceptions of confidence and self-efficacy when using the product
Personal Connection	Having a sense of belonging or personal connection to the product
Meets Expectations	Perceptions of how well the product works to meet users' needs
Product Challenges	Challenges or demands experienced when using the product
Company Empathy	Perceptions of the company in terms of trust and their intentions to design for diverse audiences

Factors in the PTPI show a strong relationship with LTR and SUS, indicating that high scores in perceptions of inclusivity may relate to a likelihood that others would recommend the product and find the product easy to use. This indicates that designing for inclusivity can have a strong effect on fundamental measures of technology product success, and measuring perceptions of inclusivity can give insights into the potential for success. Factors such as Meets Expectations and Product Challenges strongly relate to LTR and SUS. However, although scores from the Meets Expectations and Product Challenges subscales may provide insights into LTR and SUS, the PTPI can provide a more holistic assessment of the product experience with the contribution of insights from the Confidence in Usage, Personal Connection, and Company Empathy subscales.

Using the PTPI

We developed the PTPI to evaluate technology product inclusivity across any technology product. We developed the PTPI with simple language at the Grade 5 level as assessed by the Flesch-Kincaid reading ease formula (Kincaid et al., 1975) to foster comprehension across a wide range of participants. The final form of the PTPI is shown in the Appendix. When administering the scale online, it is recommended that the items be shown in a randomized order with a set of 5 items per screen. The PTPI can be scored by: (1) reverse-coding all items in the Product Challenges subscale, (2) averaging items within each subscale, and (3) summing each subscale average for a composite score. The minimum composite score is 5, and the maximum is 35.

Technology product designers can use the PTPI to assess inclusivity or to supplement other rating scales reflecting user experience (such as the SUS or LTR). Using the PTPI with a variety of technology products will allow designers and practitioners to understand the range of inclusivity by product type. For example, multi-function technology products that foster communication with friends and family, such as smartphones and smart home displays, may yield higher inclusivity scores (particularly in the Personal Connection subscale) than single-function technology products, such as an e-reader or vacuum cleaner. It would be interesting, however, to see how scores vary by brand or model to understand the influence of design on perceptions of inclusivity. The PTPI can also be used by product designers to evaluate how changes in design (such as color, customizability, shape, and style) and customizability impact perceptions of inclusivity in a single product. Additionally, the PTPI can be used to understand which areas of inclusivity are affected when users like or dislike a product. For example, lower scores in Confidence in Usage may be because users say a product is too complicated to use, whereas higher scores in Personal Connection may be because a user is able to use the same technology product as their friends or customize it to their liking.

Recommendations

We developed and validated the PTPI with a wide audience and a variety of products with the intent that the scale could be widely applicable. Although emphasis was placed on obtaining a diverse sample of different user groups (including ethnicity, age, and disability) as well as technologies, the diversity in the sample could hide trends that may be unique to a specific user group or technology category.

Limitations and Future Research

Limitations of this study include a bias for participants to choose products that they were satisfied with, which may skew the ratings. Thus, future research could focus on products that participants dislike to evaluate the model fit. Additionally, the survey software used did not distinguish between N/A responses and missing data. An N/A option was included so participants were not required to respond to a statement item if it did not apply to their technology product, so future uses of the PTPI should distinguish between missing data and N/A responses.

Future research should also be done to validate the PTPI with specific user groups and technologies. We plan to further validate and use the scale to quantify inclusion for various products, including both hardware and software, and identify areas we need to focus on to improve user experience for everyone.

Tips for Usability Practitioners

- On average, the PTPI takes approximately 5 min to administer.
- It is recommended that practitioners display 5 PTPI statements per page when administering the PTPI as a survey. This helps keep the PTPI statements from feeling overwhelming, as they may when displayed all at once.
- It is recommended to include the overall satisfaction and feelings of inclusivity questions and follow up with qualitative questions, such as "Why do you feel this item was/was not made for someone like you?"

- If asking participants to select any technology product of their choice, practitioners should include examples of technology product categories participants can choose from. This has been shown to result in a broader selection of products from participants.

Acknowledgements

The authors wish to thank Sydney Colman, Paula Conn, Blake Green, Steven Ibara, and Annie Jean-Baptiste for their assistance in this study.

References

- Ball, D., Coelho, P. S., & Vilares, M. J. (2006). Service personalization and loyalty. *Journal of Services Marketing*, 20(6), 391–403.
- Bashir, U. (2024). *Most used consumer electronics in the U.S. 2024*. Statista. <https://www.statista.com/forecasts/997113/most-used-consumer-electronics-in-the-us>
- Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25, 464–469.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Bhattacharjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems*, 19(1), 211–241.
- Bock, D. E., Mangus, S. M., & Folse, J. A. G. (2016). The road to customer loyalty paved with service customization. *Journal of Business Research*, 69(10), 3923–3932.
- Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions*, 154, 136–136.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.
- Chernev, A., Hamilton, R., & Gal, D. (2011). Competing for consumer identity: Limits to self-expression and the perils of lifestyle branding. *Journal of Marketing*, 75(3), 66–82.
- Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior*, 118, 106700.
- Chopdar, P. K., Paul, J., Korfiatis, N., & Lytras, M. D. (2022). Examining the role of consumer impulsiveness in multiple app usage behavior among mobile shoppers. *Journal of Business Research*, 140, 657–669.
- Clarkson, J., & Coleman, R. (2010). Inclusive design. *Journal of Engineering Design*, 21(2–3), 127–129.
- Clarkson, P. J., Coleman, R., Keates, S., & Lebbon, C. (2013). *Inclusive design: Design for the whole population*. Springer.
- Coelho, P. S., & Henseler, J. (2012). Creating customer loyalty through service customization. *European Journal of Marketing*, 46(3–4), 331–356.
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9.
- Davidson, S., Keebler, J., Zhang, T., Chaparro, B., Szalma, J., & Frederick, C. (2022). The development and validation of a universal enjoyment measure: The ENJOY Scale. *Current Psychology*.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–299.
- Ferdman, B. M., & Deane, B. (Eds.). (2014). *Diversity at work: The practice of inclusion*. Jossey-Bass, A Wiley Brand.
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 269–314). Information Age Publishing.

- Gaskin, J., James, M., & Lim, J. (2019). *Master validity tool* [AMOS plugin]. Gaskination's StatWiki.
- Gilal, N. G., Zhang, J., & Gilal, F. G. (2018). The four-factor model of product design: Scale development and validation. *Journal of Product & Brand Management*, 27(6), 684–700.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice-Hall, Inc.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967–988.
- Holmes, K. (2018). *Mismatch: How inclusion shapes design*. MIT Press.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Inclusion. (2022). *Cambridge Advanced Learner's Dictionary & Thesaurus*. Cambridge University Press.
- Interaction Design Foundation. (2016). *What is inclusive design?* <https://www.interaction-design.org/literature/topics/inclusive-design>
- Jean-Baptiste, A. (2020). *Building for everyone: Expand your market with design practices from Google's product inclusion team*. Wiley.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.
- Kauppinen-Räsänen, H., Björk, P., Lönnström, A., & Jauffret, M. N. (2018). How consumers' need for uniqueness, self-monitoring, and social identity affect their choices when luxury brands visually shout versus whisper. *Journal of Business Research*, 84, 72–81.
- Keates, S. (2005). *BS 7000-6:2005 Design management systems—Managing inclusive design*. British Standards Institution.
- Keates, S., & Clarkson, P. J. (2002). Countering design exclusion through inclusive design. *ACM SIGCAPH Computers and the Physically Handicapped* (73-74), 69–76.
- Keates, S., Clarkson, P. J., Harrison, L. A., & Robinson, P. (2000, November). Towards a practical inclusive design approach. In *Proceedings of the 2000 Conference on Universal Usability* (pp. 45–52). ACM.
- Kendrick, A. (2022). *Inclusive design*. Nielsen Norman Group. <https://www.nngroup.com/articles/inclusive-design/>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel*. Naval Technical Training Command.
- Lee, Y. (2020). A study on the effects of SNS use focused on social relationships, self-expression, offline activity, and life satisfaction. *Journal of the Convergence on Culture Technology*, 6(1), 301–312.
- Lee, R. M., & Robbins, S. B. (1995). Measuring belongingness: The Social Connectedness and the Social Assurance scales. *Journal of Counseling Psychology*, 42(2), 232–241.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99.
- Malone, G. P., Pillow, D. R., & Osman, A. (2012). The General Belongingness Scale (GBS): Assessing achieved belongingness. *Personality and Individual Differences*, 52(3), 311–316.
- Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, 3(1), 97–110.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

- O'Brien, H. L., & Toms, E. G. (2009). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1), 50–69.
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined User Engagement Scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28–39.
- Park, C. S., & Kaye, B. K. (2019). Smartphone and self-extension: Functionally, anthropomorphically, and ontologically extending self via the smartphone. *Mobile Media & Communication*, 7(2), 215–231.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47(3), 537–560.
- Rowlands, L. (2015). "Ugly" hearing aid ad leaves parents fuming. *Stuff*. Retrieved September 12, 2022, from <https://www.stuff.co.nz/life-style/parenting/baby/caring-for-baby/68919793/ugly-hearing-aid-ad-leaves-parents-fuming>
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin*, 28, 1629–1646.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sashi, C. M. (2012). Customer engagement, buyer–seller relationships, and social media. *Management Decision*, 50(2), 253–272.
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1), 3–15.
- Selig, A. (2020). Generation influence: Reaching Gen Z in the new digital paradigm. *WP Engine*. Retrieved September 12, 2022, from <https://wpengine.com/resources/gen-z-2020-full-report/>
- Sheldon, K. M., & Hilpert, J. C. (2012). The Balanced Measure of Psychological Needs (BMPN) scale. *Motivation and Emotion*, 36, 439–451.
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79(2), 310–334.
- Shin, S. A., Jang, J. O., Kim, J. K., & Cho, E. H. (2021). Relations of conspicuous consumption tendency, self-expression satisfaction, and SNS use satisfaction of Gen Z. *International Journal of Environmental Research and Public Health*, 18(22), 11979.
- Steiger, J. H. (1980, May). Statistically based tests for the number of common factors. *Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA*.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The sense of agency scale. *Frontiers in Psychology*, 8, Article 1552.
- Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management*, 24(1), 53–62.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.
- Vagias, W. M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838.

Appendix

PTPI

Instructions: Please rate the following statements on a scale from "Strongly disagree" to "Strongly agree." If a statement does not apply, select "N/A" (see Figure).

Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree	⊗ N/A
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Personal Connection

1. I have a sense of belonging when I use this product.
2. The look of this product allows me to feel like I belong.
3. I feel a personal connection to this product.
4. When using this product, I feel my choices express my "true self."

Product Challenges

5. This product is emotionally demanding to use.
6. This product is mentally demanding to use.
7. This product is physically demanding to use.
8. For this product to work, I had to make changes to it beyond my expectations.
9. It's hard for me to use this product on my own.
10. When using this product, I struggle to do things I should be good at.
11. When using this product, I feel like my actions had unintended consequences.

Confidence in Usage

12. I am confident that I know how to use this product.
13. It is easy for me to learn how to use this product.
14. I am good at using this product.
15. I feel very capable using this product.
16. It's easy for me to remember how to use this product.

Meets Expectations

17. This product meets my expectations.
18. This product is reliable.
19. I consider my product usage experience a success.
20. This product works well for me.
21. I feel in control of my product experience.
22. There is a good fit between what this product offers me and what I am looking for in this product.

Company Empathy

23. Overall, the company that made this product is trustworthy.
24. The company that made this product makes good-faith efforts to address the concerns of customers like me.
25. I feel like the company considered the needs of customers like me when designing this product.

About the Authors



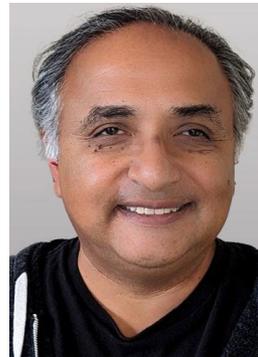
Carmen Van Ommen
Carmen Van Ommen is a doctoral candidate at Embry-Riddle Aeronautical University. Her research focuses on accessibility, gaming, and usability. She works within the Research in User eXperience (RUX) Lab, and her recent work includes evaluating video game satisfaction for people with disabilities and the user experience of adaptive game controllers.



Barbara S. Chaparro, Ph.D.
Dr. Barbara S. Chaparro is a Professor at Embry-Riddle Aeronautical University and head of the Research in User eXperience (RUX) Lab. Her research interests include the study of factors that influence the user experience (UX) of products, software and systems, the investigation of usability assessment methods, and the efficacy of augmented/mixed reality devices and applications.



Joseph R. Keebler, Ph.D.
Dr. Joseph Keebler, PhD, is a professor of Human Factors at Embry-Riddle Aeronautical University and is the director of the Small Teams Analog Research (STAR) lab. His research focuses on training, simulation, and teamwork within high-risk domains like aviation and medicine.



Sanjay Batra, Ph.D.
Sanjay Batra is a UX Research Manager and Accessibility Specialist at Google, with over 25 years of experience in consumer hardware and software research. His work centers on crafting inclusive products and experiences. He puts inclusion at the core of team culture, people, and processes, and is actively involved in inclusive hardware design for Pixel and Nest products.



Mei Lu
Mei Lu has over 20 years of experience in User Experience research. She is currently a UX Research Lead at Google, working on the Design for All team.