



Development of the Empathy in Design Scale: Measuring Employees' Empathy Toward Users in Service Design

Luce Drouet [†], Kerstin Bongard-Blanchy and Carine Lallemand ^{†,*}

Maison des Sciences Humaines, University of Luxembourg, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg

*Corresponding author: carine.lallemand@uni.lu

[†]The first and last author have equally contributed to this work.

Abstract

To design user-centered services, it is essential to build empathy toward users. It is hence strategic to trigger empathy for users among professionals concerned with shaping service user experiences. There is, however, a lack of quantitative tools to measure empathy in design. Through two studies, we report on the development and validation of the Empathy in Design Scale (EMPA-D). The tool aims to measure service employees' empathy toward users. Grounded in theories from psychology and design, we first generated and tested a pool of items through expert inspection and cognitive interviews. In Study 1, we administered 16 items to 406 full-time service employees from various industries, including employees in customer-facing positions. In Study 2, we iterated on additional items and administered a revised scale to 305 service employees. The selected model consists of 11 items and has a three-factor structure (*Emotional interest/Perspective-taking*, *Personal experience* and *Self-awareness*), which showed an adequate model fit and good internal consistency. Evidence of convergent validity was provided by moderate correlations of the EMPA-D scale with empathy measures in psychology (SITES, Empathy Quotient, Interpersonal Reactivity Index), whereas discriminant validity was demonstrated by low correlations with the narcissism measure Narcissistic Personal Inventory. We outline how this self-reported empathy measure can support organizations in enhancing their services and discuss potential limitations of quantitatively measuring empathy in service teams.

RESEARCH HIGHLIGHTS

- We present the development and validation of the Empathy in Design Scale (EMPA-D), a self-report measure of employees' empathy toward users of a service.
- We report on two validation studies and document the psychometric properties of the scale.
- The selected model consists of 11 items and a three-factor structure (*Emotional interest/Perspective-taking*, *Personal experience* and *Self-awareness*).
- The resulting EMPA-D scale contributes to filling the gap in metrics to assess empathy in the service design context.
- In industry, measuring employees' empathy support the selection of appropriate empathic interventions to foster the service user-centeredness.

Keywords: *empathy; empathic design; scale development; measurement tool; service design; method*

1. INTRODUCTION

Users expect high-quality experiences from products and services. To deliver such experiences, designers require a solid understanding of user expectations, needs and values. Over the last decades, empathy has gained importance in design (Koskinen *et al.*, 2003; Sleeswijk Visser and Kouprie, 2008; Dong *et al.*, 2018)—as evidenced by the emergence of the Empathic Design approach and methods in the 1990s (Leonard and Rayport, 1997; Koskinen *et al.*, 2003). Empathy allows designers to put themselves into the users' shoes (Lietz *et al.*, 2011), feel more engaged with the users and better understand their perspectives (Devecchi and Guerrini, 2017). Service companies increasingly seek to develop empathy inside their organization and toward customers, as it increases customers' satisfaction, loyalty and forgiveness toward

the service (Wieseke *et al.*, 2012; Bove, 2019; Bahadur *et al.*, 2020). Empathy is a factor of service employee performance (Nguyen *et al.*, 2019; Ngo *et al.*, 2020), which guarantees service quality and fosters innovation (McDonagh, 2010; Bove, 2019).

At the organizational level, empathy is the core of culture transformation toward user-centered approaches. The 'awareness' stage of the user experience (UX) maturity model implies making 'every employee aware of their responsibility towards users and their experiences with the service' (Chapman and Plewes, 2014). Services have the particularity to be co-created by service designers (Stickdorn *et al.*, 2018) and other service employees (Roto *et al.*, 2020; Lee *et al.*, 2022a), whose mission tends to be more solution-oriented rather than user-oriented. In this paper, we focus on employees as service stakeholders: persons involved

Received: May 31, 2023. Revised: May 2, 2024. Accepted: May 13, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of The British Computer Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

in creating or providing a service who operate within the bounds of their organization, e.g. frontline employees, maintenance operators, communication officers and IT engineers (Stickdorn et al., 2018). Involved in shaping the service, their empathy toward users plays a crucial role in the experience of users: ‘If the people within the organization do not share the attitude or mindset that is needed for doing empathic design, then the effort is likely to strand’ (Postma et al., 2012).

While the importance of building empathy in design is undisputed, little is known about how it can be actively fostered. Being able to measure empathy quantitatively would be a first step to better understand these mechanisms (Chang-Arana et al., 2022). To the best of our knowledge, no quantitative instruments are currently available to measure empathy within the service or product design context (Surma-aho and Hölttä-Otto, 2022). Besides designers, other service stakeholders have been somewhat overlooked in the design empathy literature so far. Measuring employee’s empathy can support the selection of appropriate empathic interventions (Postma et al., 2012; Hess et al., 2017; Drouet et al., 2024) to deploy within the company. Recent initiatives within the design community have emphasized the need and relevance of empathy measures (Chang-Arana et al., 2022), whether self-reported *empathic tendencies* (Surma-aho and Hölttä-Otto, 2022) or the assessment of the outcomes resulting from empathic interventions as evidence of increased empathy (Sleeswijk Visser and Kouprie, 2008; Drouet et al., 2024).

Drawing on frameworks and methods for measuring empathy, particularly in the design context, this paper presents the steps undertaken to build and validate an empathy scale for measuring service employees’ empathy. EMPA-D methodologically contributes to the service design field by providing a novel quantitative self-report measure. The tool is cost-efficient and suitable for research and industry alike. In industry, the scale aims to support organizations by assessing the employees’ empathy toward users. In academia, such a measurement tool is relevant to support conceptual discussions as it allows modeling and predicting variables (Chang-Arana et al., 2022), such as the influence of empathy on the success of design interventions.

2. RELATED WORK

2.1. Definitions of empathy in design

There is no consensus on the definition of empathy in design, which is often referred to as an ‘umbrella term’ (Chang-Arana et al., 2022). Design research builds its understanding of empathy on philosophy, psychology and neurosciences, which follow distinct epistemological approaches (Surma-aho and Hölttä-Otto, 2022). Cuff et al. (2016)’s review illustrates the complexity of defining empathy, with no less than 43 definitions of empathy in psychology. To scope the concept, the authors underline the distinctions between:

- empathy and other concepts such as sympathy or compassion
- cognitive empathy—‘the ability to ascribe mental states to others, such as beliefs, intentions, or emotions’ (Maibom, 2017)—and affective empathy involving ‘affects on the part of the empathizer.’ Both forms of empathy are related and interdependent (Hodges and Biswas-Diener, 2007).
- self/other (Chang-Arana et al., 2022): the empathizer does not experience others’ emotions but takes their viewpoint and is aware that others’ emotional responses are distinct from their own.
- empathy as a trait of personality—some individuals are more empathic than others, or a state—meaning that being empathic depends on the context (Shen, 2010; Cuff et al., 2016), or even a skill (Battaly, 2011; Walther et al., 2017). According to Hodges and Biswas-Diener (2007), empathy is a state of mind that people can control and modify. Likewise, empathy does not trigger a direct behavioral response (e.g. prosocial behavior) but rather the behavioral motivation to act (Cuff et al., 2016). This motivation is key in the design literature (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b; Smeenk et al., 2018, 2019).

According to Tellez and Gonzalez-Tobon (2019), empathy is the ‘ability to feel and understand other’s emotions and circumstances and is a fundamental skill for designers to acquire an in-depth understanding of people (i.e., end-users and other stakeholders) so that products, services, environments, systems, and experiences meet human needs, expectations, and aspirations.’ Fulton Suri (2003) describes it as the designers’ navigation between two conceptions of others’ experience: (i) thinking that others are like oneself and will obviously understand what one has designed and (ii) thinking that others are completely different from oneself.

2.2. Frameworks of empathy in design

Besides defining the concept of empathy, researchers have investigated the process of empathy in design. Some early publications (e.g. Leonard and Rayport, 1997; Koskinen et al., 2003) address empathic design as a user research process, adopting a methodological or practical approach that does not inform the conceptualization of empathy. Fulton Suri (2003) describes it as the designers’ navigation between two extreme conceptions of others’ experience: (i) thinking that others are like oneself and will understand what one has designed and (ii) thinking that others are completely different from oneself. Imagination plays a crucial role in this navigation. It is fed by reality and needs to be continuously checked with reality (Fulton Suri, 2003). Building on these early work, design researchers iteratively formalized empathy in design frameworks (Chang-Arana et al., 2022), which constitute the foundation of the present contribution (Fig. 1). Preston (Preston, 2007; Preston and De Waal, 2002) inspired theories of empathy in design by establishing that an empathizer must be motivated to and capable of behaving and responding empathically in three ways: attune, experience and respond. For Smeenk et al. (2019), this means that designers: (i) should consciously attend and attune to the state of the users, (ii) must be willing to open themselves up to experiencing a similar emotional state and/or to activating similar personal experiences, and (iii) should generate a suitable and sensitive emotional response to users.

Kouprie and Sleeswijk Visser (2009) introduced a four-phase framework for empathic design based on the work of psychologists like Stein et al. (2017) and Rogers (1959). The model illustrates a deliberate act of stepping in and out of others’ lives, appealing to both cognitive (gray in Fig. 1) and affective empathy (in white in Fig. 1). First, designers approach the users’ world through discovery. This triggers their curiosity and willingness to understand users’ experiences. Then follows the immersion, where they internalize the users’ point of reference without judging them. During the connection phase, they reflect on their own experiences. In the final detachment phase, they regain distance to find solutions and ideate.

Other authors defined the landscape of empathy in design through matrixes or intersecting dimensions, initially 2D (Sanders and Stappers, 2008; Sanders and Stappers, 2012) and later 3D

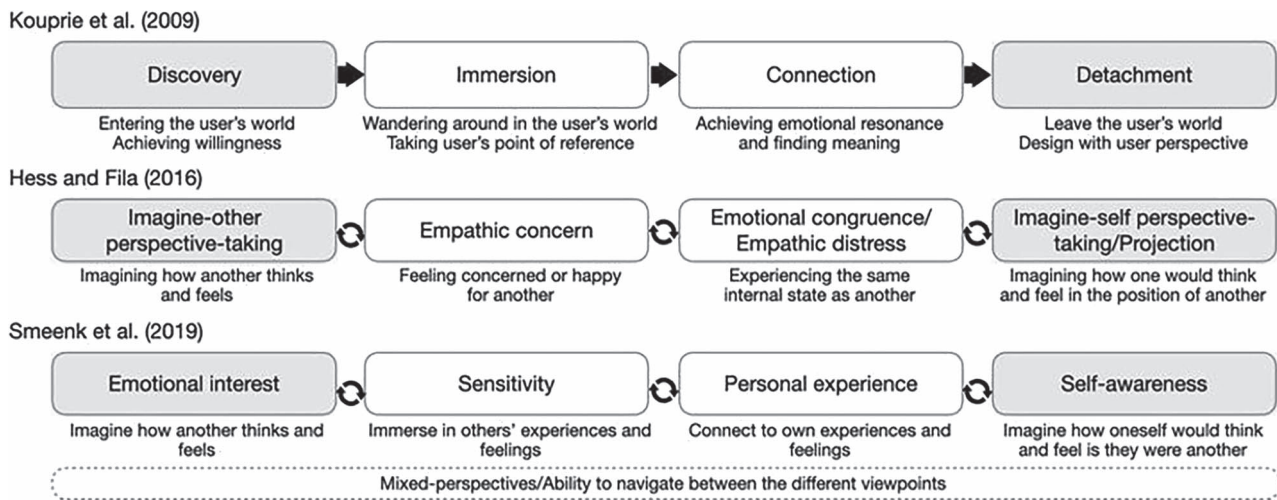


Figure 1. Synthesis of the main frameworks of empathy in design and their overlaps

like the system of coordinates on empathy proposed by [Dong et al. \(2018\)](#). Two dimensions are commonly found in models of empathy in design (with various labels): the opposition versus self and others and the distinction between cognitive and affective empathy ([Hess and Fila, 2016a](#), [Dong et al., 2018](#), [Smeenk et al., 2018](#), [Sandman et al., 2020](#)). Their combination results, for [Hess and Fila \(2016b, 2016a\)](#), in four dimensions: (a) *imagine-other perspective-taking* when designers imagine how the users think and feel, (b) *empathic concern* when designers feel concerned for the users, (c) *emotional congruence* when designers experience the same emotional state as users, and (d) *imagine-self perspective-taking/projection* when designers imagine how they would think and feel if they were the users. Based on this taxonomy, [Smeenk et al. \(2018, 2019\)](#) suggest the notion of empathic formation—“the formative process of becoming an empathic design professional”—with the aim to identify the factors that foster empathy in design from a more contextual position. Their Empathic Formation Compass, integrating most dimensions identified in prior work, includes the three perspectives a designer can take (the first-, second- and third-person perspective). It also represents the information sources and the factors that foster empathy in design. Each perspective relates to a specific source of information, which, in turn, stimulates a distinct mindset: *emotional interest*, *sensitive*, *personal experience*, and *self-aware*. Designers’ *emotional interest* is when they choose to be receptive to users’ experiences. The *sensitivity* develops when the designers are in contact with the users. *Self-awareness* refers to designers’ ability to distinguish their own experiences from that of the users. When designers are familiar with the use context, they reflect on their *personal experience*. An additional *mixed-perspective* refers to the ability to navigate between these viewpoints.

These frameworks largely overlap in their conceptualization of empathy in design ([Fig. 1](#)), with similar content areas adopting slightly different labels or overarching categories. Let us consider taking the user’s point of reference as an illustration: [Hess and Fila \(2016a, 2016b\)](#)’s *Perspective-taking* dimension corresponds to the *Discovery/Emotional interest* stage of other frameworks, but [Kouprie and Sleeswijk Visser \(2009\)](#) place it in the *Immersion* stage. Some concepts are more controversial, like the idea of empathic distress included in [Hess and Fila \(2016a\)](#)’s model but described by [Preston \(2007\)](#) as something to “inhibit.”

Overall, these frameworks offer worthwhile foundations to develop tools for measuring empathy in design. What is now

needed to consolidate theoretical claims are measurement tools operationalizing these dimensions to support the collection of empirical evidence through experimental research ([Surma-aho and Hölttä-Otto, 2022](#)).

2.3. Value of empathy in service design

Service design is a holistic approach focused on end-user experiences, as well as service stakeholders and suppliers’ experiences involved in shaping services ([Stickdom et al., 2018](#); [Roto et al., 2020](#)). It is concerned with customers, providers and stakeholders’ journey and experience ([Roto et al., 2020](#)). The question of the value is central to understanding the services. In contrast with the value-in-exchange of product logic, service logic focuses on value-in-use ([Ng et al., 2009](#)). From people’s perspective, services deliver value in use through interactions between users and the service company ([Vargo et al., 2008](#)). A variety of stakeholders are involved in shaping and delivering services, and collaboration with customers is a key aspect ([Kuusisto and Päälysaho, 2008](#)).

According to [Tripp \(2013\)](#), there is no service without empathy. Beyond designers, in a service context, empathy sparks service employees’ awareness of their role in users’ experiences. This awareness is the ground for increasing organizations’ user experience maturity ([Chapman and Plewes, 2014](#)). On the customer side, empathy develops their loyalty and forgiveness toward service companies resulting in better customer satisfaction ([Marandi and Harris, 2010](#); [Wieseke et al., 2012](#); [Bahadur et al., 2020](#)). Recent research studied the importance of frontline staff empathy ([Nguyen et al., 2019](#)), advising service companies to recruit frontline staff motivated to understand customers ([Saco and Goncalves, 2008](#)). Empathic employees recognize customers’ needs more easily than others, improving human interactions and the employee–customer relationship ([Bove, 2019](#)). Empathy increases employees’ attentiveness toward customers and performance ([Wieseke et al., 2012](#); [Nguyen et al., 2019](#); [Ngo et al., 2020](#)). In return, it builds customers’ empathy toward the service and its staff ([Wieseke et al., 2012](#)).

2.4. Empathy scales in psychology

Various scales measuring empathy have been developed in psychology or related disciplines. Among them, the popular Empathy Quotient (60-item) ([Baron-Cohen, 2004](#)), the Interpersonal Reactivity Index ([Davis, 1980](#)), the Toronto Empathy Questionnaire ([Spreng et al., 2009](#)), the Empathy Assessment

Index (Lietz *et al.*, 2011) or Hogan's (1969) scale. Many of these scales include an abbreviated version, e.g. Empathy Quotient (Baron-Cohen, 2004) and Empathy Assessment Index (Lietz *et al.*, 2011). A single-item Likert-scale of empathy SITES has even been proposed by Konrath *et al.* (2018): 'I am an empathetic person (Note: An empathetic person understands others' feelings, and experiences care and concern for them)'. Most of the multidimensional scales used in psychodiagnostics assess empathy competencies or responses, for instance, in the context of behavioral disorders in early childhood (Cuff *et al.*, 2016). Other examples, like the Jefferson Scale of Empathy (Hojat, 2016), measure empathy among healthcare professionals. These tools are generally based on statements assessing several facets of empathy. The ratings are made via Likert scales (Davis, 1980; Baron-Cohen and Wheelwright, 2004), behavior description or dichotomous Yes/No choices (Stadler *et al.*, 2004). Scenario-based tools asking respondents to infer characters' emotional state are also used (e.g. Petermann and Gust, 2016). Surma-aho and Hölttä-Otto (2022) distinguish six indicators of empathy that are assessed in psychology and related sciences: (i) *empathic tendencies*, referring to self-reported tendencies to feel or understand how the other is feeling in everyday situations (Spreng *et al.*, 2009). These are measured in surveys by asking how a person would react or feel in given scenarios, (ii) *beliefs about empathy*—are about how a person values empathy or related issues. These self-reported measures capture how important empathy is for the respondent. (iii) *Emotion recognition*, the ability to identify the emotion(s) another person is demonstrating or experiencing, is commonly assessed using behavioral and performance-based instruments, (iv) *understanding mental contents*—which goes deeper than emotion recognition by also guessing others' thoughts, (v) *shared feelings*—the synchronicity between others' emotions and oneself (emotional empathy), measured in neurosciences as similarity of some physiological signals and (vi) *prosocial responding*, which uses reactions or behaviors (e.g. willingness to help) as indicators of empathy.

In psychology, empathy scales cover the measure of the state and the trait of personality while design research is primarily concerned with the state. Empathy is a state of mind in design that serves the objective of designing solutions adapted to users' reality. Design measure tools are needed to assess the aspect of empathy allowing to achieve this goal.

2.5. Empathy scales in service research

In service research, empathy measures are carried out from the customer's point of view and focus on service quality like with the tool SERVQUAL (Parasuraman *et al.*, 1988). This questionnaire includes 22 items assessing five dimensions using a 7-point Likert-scale. Empathy is defined as '*caring, individualized attention the firm provides its customers*'. Five items address customer perception of employees' empathy toward them, e.g. 'Employees of XYZ do not give you personal attention' or 'Employees of XYZ do not know what your needs are'. SERVQUAL has been extensively used for the last decades but has also been criticized for theoretical and operational flaws (Buttle, 1996). Despite its shortcomings, it is one of the first tools including the employees' attitude when reflecting on empathy toward users. This last decade, service research focused on researching empathy's impact on customer satisfaction, loyalty and forgiveness toward the service (Bove, 2019). Rare service research measures empathy to explain its impact on service design. Wieseke *et al.* (2012) measured employees' empathy (i.e. *Perspective-taking*, *Empathic concern* and *Emotional contagion*) to assess the role of empathy in customer–employee interactions,

through McBane (1995)'s measurement tool, and Bahadur *et al.* (2020) assessed the effect of employee empathy on service loyalty with four generic items on employee empathy.

2.6. Empathy assessment techniques in design

A few studies measure designers' empathy during the design process. Given the lack of a validated self-report scale specific to the domain (Surma-aho and Hölttä-Otto, 2022), design researchers resort to alternative approaches. The observation of prosocial behavior (e.g. helping someone, taking altruistic decisions) can inform about empathy (Surma-aho and Hölttä-Otto, 2022). Discourse analysis supported by qualitative data analysis techniques (Weber *et al.*, 2022; Lee *et al.*, 2022b) or machine learning (Wambsganss *et al.*, 2022, 2021), can reveal facial or vocal expressions of empathy (Salmi *et al.*, 2022), or pro-social language (Rojas *et al.*, 2022). For instance, van Rijn *et al.* (2011) analyzed the discourse of designers during teamwork, looking for four empathy indicators: empathic expressions (e.g. saying 'I think/feel/guess the users think/feel/want...'), own experience (e.g. relating users' needs and experiences to their personal experiences or comparing them to people they know), questioning users' needs and experiences vs. making (false) assumptions and discussing user facts. Sleeswijk Visser and Kouprie (2008) tested their framework for empathy design in a workshop with design practitioners. To measure the effectiveness of their empathic user research intervention, they combined the Empathy Quotient (Baron-Cohen, 2004) with the observation of specific verbal expressions. They also used *ad hoc* scales through which designers self-reported their interest, involvement, inspiration and empathy at regular intervals during the workshop. Finally, they recorded the evolution of the designers' empathy using experience curves. Finally, Chang-Arana *et al.* (2020) integrated the empathic accuracy method into the design process. It consists of successive activities: interviews between designers and users are video-recorded. The user watches the recording and annotates key moments by describing their thoughts. Up to 4 weeks later, the designer is shown the video and infers what the user might have thought at each annotated point. The researchers assess the similarity between the inferred and actual thoughts on a three-point scale. While the empathic accuracy approach can assess designers' empathy to a certain degree, it is complex and time-consuming. Responding to this critique, the Quick Empathic Accuracy involves a shortened experimental protocol with fewer entries (Li *et al.*, 2021). Closely related to design practice is the measure of engineers' empathy with the Empathy and Care Questionnaire (Hess *et al.*, 2017).

The assessment of empathy in design is mainly based on qualitative metrics. Existing studies using quantitative measures of empathy are scarce in the design domain and typically use psychological instruments. However, these instruments are not always suitable for design purposes as they tend to measure empathy as a personality trait rather than a state and thus do not align with empathy in design frameworks. Away from clinical outlooks, a state-based measure of empathy in service design would seek to indicate which empathy dimensions need to be triggered to lead to a shared commitment of employees toward users. Notwithstanding the relevance of qualitative studies, we concur with recent work emphasizing that empathy in design would benefit from quantitative approaches supporting robust hypothesis testing (Chang-Arana *et al.*, 2022). The development of specific empathy in design scales would thus support researchers in testing and refining current theoretical models.

2.7. Scale development

Scale development is a foundational process in psychology and social science research for quantifying abstract constructs and facilitating empirical investigation. Numerous guidelines have been formulated for conceptualizing, constructing and evaluating scales, from the classical *Psychometric theory* by Nunnally (1978), to Spector's (1992) *Summated rating scale construction* and more recent contributions (e.g. Kline, 2000; DeVellis, 2017; Furr and Bacharach, 2013; Kyriazos and Stalikas, 2018).

Scale development entails various stages and follows key considerations, detailed in the aforementioned literature. In a nutshell, scale development starts with the clear conceptualization of the target construct (in our case, empathy in design). To generate an initial pool of items, rigorous item generation techniques are employed, including a thorough literature review, experts' involvement and qualitative research methods. Subsequent item selection and refinement processes ensure that selected items demonstrate relevance, clarity and psychometric properties. Selecting appropriate response formats (e.g. Likert scales, semantic differentials or visual analogue scales) is key for capturing nuanced responses from respondents. Scales are usually pilot tested with a small sample to assess the clarity of items, their comprehension and appropriateness before the scale is administered to a large sample of respondents. This stage uses psychometric methods such as factor analysis and structural equation modeling to enhance scale refinement and validation.

Scale development guidelines detail fundamental psychometric properties to evaluate in order to ascertain the robustness of the scale, namely, reliability, validity and sensitivity. Validity refers to the extent to which the scale accurately measures the intended construct. It encompasses aspects such as content validity, criterion validity and construct validity. Reliability measures such as internal consistency and test-retest reliability assess the consistency and stability of scale scores. Sensitivity analysis evaluates the scale's ability to detect meaningful differences in the construct of interest, crucial for longitudinal or intervention studies. Rigorous scale development practices are essential for producing reliable and valid measurement tools.

3. RESEARCH OBJECTIVES

While qualitative studies dominate research on empathy in design, a quantitative approach 'allows predicting, controlling and explaining, [and is] needed to test the causality of empathy in successful design' (Chang-Arana et al., 2022). Existing work, furthermore, majorly focuses on the empathy of designers toward users rather than addressing the empathy of the larger group of employees who provide and shape services. The development of the Empathy in Design Scale (EMPA-D) aims at providing a validated quantitative indicator of service employees' empathy toward users.

As a self-reported measure, EMPA-D addresses the *tendency* to be empathic and underlying *beliefs* (Surma-aho and Hölttä-Otto, 2022). The creation of such a tool contributes to filling the gap in metrics to assess empathy in the context of service design as there is, to the best of our knowledge, no measurement scale of employees' *empathic tendency* (Smeenk, 2019; Chang-Arana et al., 2022; Surma-aho and Hölttä-Otto, 2022) toward users. For industry, the scale aims to support organizations in their growth toward user-centeredness by enabling them to assess their employees' empathy. Quantifying dimensions of empathy will point to shortcomings and guide intervention strategies (i.e. using a specific

empathic method). For academia, the scale will provide a quantitative measure of empathy, relevant to investigate and model several phenomena related to the design process (e.g. influence of empathy on the success of design interventions). The data underlying this article will be shared on reasonable request to the corresponding author.

4. EMPA-D INITIAL DEVELOPMENT PROCESS

4.1. Definition and scope of the construct

Following the best practices on summated rating scale construction (Spector, 1992; DeVellis, 2003), we first conducted a literature review to define the content areas relevant to the evaluation of empathy in a design context. Building on the frameworks of empathy in design by Kouprie and Sleeswijk Visser (2009), Hess and Fila (2016b) and Smeenk et al. (2018, 2019), the *empathic tendency* includes an initial *discovery*, an *immersion* into the user's world, a *connection* with one's own *personal experience* and a *detachment* from the users' viewpoint. For designers, the ability to navigate across these steps, and the underlying cognitive or affective empathy types, is key (*mixed-perspective* in Smeenk et al., 2019). From these frameworks, we synthesized and merged similar definitions of empathy and we here forward refer to empathy in design as the ability of designers and service stakeholders to understand users and their experiences, which is reflected by an emotional interest toward users, the sensitivity and ability to actively take the user perspective, build on their own personal experiences to connect with users and keep a self-awareness of others' reality. Empathy in design motivates action and engagement. We envision empathy in design as a multi-dimensional construct, based on four potential content areas (Table 1): *Emotional interest/Discovery* (EI), *Sensitivity/Immersion* (S), *Personal experience/Connection* (PE), and *Self-awareness/Detachment* (SA). These content areas represent the specific aspects or domains of the empathy in design construct that our scale aims to measure. The empirical dimensionality of the scale will be determined through factor analysis in the subsequent steps.

4.2. Creation of a pool of items

Two of the authors independently generated a pool of items for each of the four content areas based on a review of self-reported instruments of empathy in psychology, specific design frameworks and empirical studies. This resulted in six to eight items per content area (33 in total), with a natural overlap yet subtle variance between certain items. The authors then merged or adjusted items showing the most redundancy, resulting in 29 items. All statements referred to service design but could be adapted to system or product design, as the most frequent wording used in the items revolves around the experience of the users.

When designing a measurement scale, the number and type of response categories were carefully considered, as they may influence the psychometric data quality. Inspired by Kashdan et al. (2020) and following standards in our field, a 7-point Likert-scale to indicate 'the degree to which these statements accurately describe you or not' was chosen and successfully pilot tested for comprehension and sensitivity. It offers a high granularity inviting respondents to make less polarized choices (Simms et al., 2019) and allows participants to adopt a neutral position. Despite debates on the ambiguity of using the middle option on odd-numbered response scales (Kulas and Stachowski, 2013), it seems worthwhile to leave the opportunity to indicate neutrality in

Table 1. Description of the four content areas of empathy in design in the construction of the Empathy in Design Scale (based on Hess and Fila, 2016b; Kouprie and Sleeswijk Visser, 2009; Rogers, 1959)

Content area	Definition
Emotional interest/Discovery (EI)	Imagining how users think and feel—curiosity about the users, resulting in a willingness and motivation to explore and discover the users, their situations and experiences.
Sensitivity/Immersion (S)	Direct contact—taking an active role by wandering around in the user's world, absorbing without judging. Being open-minded and interested in the user's point of reference.
Personal experience/Connection (PE)	Resonating with the user—connecting with users on an emotional level, by recalling explicitly upon one's own memories and experiences in order to reflect and be able to create an understanding.
Self-awareness/Detachment (SA)	Stepping back in the role of designer or stakeholder—correctly distinguishing between the representations of one's own actions, perceptions, sensations, and emotions and those of users (Smeenk et al., 2019).

our context. All statements were positively formulated to avoid additional cognitive load, and potential errors linked to the use of double negatives (Swain et al., 2008; Weijters and Baumgartner, 2012), potentially impacting the internal consistency and factor structure of a scale (Zeng et al., 2020). We closely monitored the responses during pre-testing to detect potential acquiescence or response biases. We included attention checks in the administration of the scale to prevent carelessness or non-contingent responses.

4.3. Initial expert evaluation

We conducted an expert review of the pool of items, following Cabrera-Nguyen (2010)'s scale development guidelines. The goal was to validate whether the items measure empathy and are related to the content area to which they are attributed. We submitted the pool of 29 items to three experts with cognitive psychology and interaction design backgrounds and knowledgeable in empathy, service design and scale development. We checked the scale's face validity by asking them to review each item separately and categorize it in the predefined content areas. They could comment both on their understanding of the item and on language issues. This resulted in a reformulation of items and a decrease from 29 to 17. Some items did not obtain consensus in terms of face validity or clarity; others were redundant in the content area measured. In the following stages, we kept redundant items to identify the most suitable ones.

4.4. Initial version and pretest

We pilot-tested the initial version to ensure item understandability and detect potential response bias. As the models and frameworks in design used in the creation of the item pools mostly applied to designers, we can wonder whether the same dimensions can relate to the empathic tendencies of service employees delivering and partially shaping the service. We pretested the 17-item version of EMPA-D with $n = 8$ digital service employees (five men, three women, aged 23–45, fluent English speakers). We used cognitive interviewing (Castillo-Díaz and Padilla, 2013) to investigate how participants interpreted and responded to each item. Beyond ensuring the items understandability, this technique constitutes a source of construct validity evidence (DeVellis, 2003; Castillo-Díaz and Padilla, 2013). Each participant was invited to think aloud while filling the scale, commenting on the way they interpreted each statement and on the rationale behind each of their answer. Participants were not informed on the purpose of the scale, which was presented as a questionnaire about 'working in a service context'. Additionally, participants were also asked to rate the understandability of each item on a 5-point Likert-scale (1—'Not understandable at all', 5—'Totally understandable') and to

explain their score. We analyzed their item ratings, explanations and understandability ratings.

We noted some comprehension issues of key terms or errors or ambiguity in the item wordings. While 5 out of 17 items reached very high understandability scores of 4 or 5 (EI1, EI3, EI4, SA2, SA4), the variance was rather high for the remaining 12 items. The average understandability score was nevertheless high for seven other items with a mean score above 4 (EI2, S1, S2, S4, SA3, PE2, PE4). The items with the lowest understandability score were S5 ($M = 3.25$, $SD = 1.67$), PE1 ($M = 3.25$, $SD = 1.75$) and SA1 ($M = 3$, $SD = 1.51$). Regarding S5, participants raised concerns that not all service employees can easily 'go to the field', while customer-facing staff is already there. As the term was considered jargon, we removed this item. Item SA1 originally read—'I imagine how I would think and feel if I were a regular user rather than an employee'. Here, 'regular' aimed at distinguishing users from the respondent when the employees too are potential users of the service. This was not well understood and therefore removed. The absence of context and mention of frequency in item PE1—'I often consider and reflect on my own experiences and feelings'. led to confusion. The reformulated item omits 'often' and clarifies the context. Item S2—'I pay attention, without judging, to how users experience the service' was understood ($M = 4.25$, $SD = 0.89$) yet had a complex construction and was simplified.

We contextualized and refined the instructions to refer more explicitly to the participant's professional context. In addition, the answer scale option 'Neutral' (used in Kashdan et al., 2018's social curiosity scale) was perceived as unsuitable by some respondents and reformulated into another common Likert anchor label 'Moderately describes me' in line with the surrounding items. Finally, we looked at the distribution of pretest answers on the scale to explore the scale sensitivity (i.e. capacity to discriminate between individuals with different levels of empathy). Several items showed a skewed tendency with high agreement scores, especially in the self-awareness dimension. It was thus a point of attention in future steps to ensure that common biases such as acquiescence or social desirability (Krumpal, 2013) would not affect the psychometric properties of the tool. Following the recommendations by DeVellis (2017) about the formulation of statements used in Likert scales, we thus avoided too mild statements that elicited too much agreement. A particular attention was paid to a balance the strength of a statement for sensitivity purposes with social desirability concerns, which are exacerbated in the envisioned context of use (a company) of an empathy scale targeting service employees. Overall, this stage resulted in a 16-item experimental version.

5. STUDY 1: ADMINISTRATION OF THE EXPERIMENTAL VERSION

5.1. Participants

We administered the EMPA-D to $n = 406$ valid respondents (212 men, 193 women, 1 undisclosed). Eight respondents were excluded from the sample due to failing the attention checks. Participants' age ranged from 19 to 76 ($M = 37.4$, $SD = 11.07$). We used the crowd-sourcing platform Prolific.co. The participants were native English speakers and full-time employees. Building on the Prolific sampling options for professions in the service sector, we administered the scale to service employees from various industries, e.g. public administration, hospitality/tourism, transportation and logistics, information technology, marketing/retail ($n = 226$) and employees in customer-facing positions ($n = 180$). All respondents indicated to what extent (from 1 to 7) they work in direct contact with users ($M = 5.7$, $SD = 1.09$). The time spent in their current organization ranged from <1 to 47 years ($M = 7.3$, $SD = 8$). Over half of the participants (58%) held a university degree. The study was approved by the University of Luxembourg ethics panel. Participants gave informed consent and received fair compensation. The study was labeled 'Working in a service company' to not prime respondents on empathy.

5.2. Material and procedure

The study was created on the Limesurvey platform, accessed by the participants via a URL on the Prolific.co crowdsourcing platform. Following the demographic questions, we administered the experimental version of EMPA-D composed of 16 items, presented in a randomized order. Each item is a first-person statement rated on a 7-point Likert-scale, from 1- 'Does not describe me at all' to 7- 'Completely describes me' (Appendix A). The statements were presented in the form of a table, on a single page, with alternate row shading to avoid errors. The instruction read: 'The table below includes statements related to your professional context. By "users" we mean people/customers who use your organisation's products or services. Please use the 7-points scale to indicate the degree to which these statements accurately describe you or not. Respond spontaneously: there are no right or wrong answers, only your perspective matters.'

5.3. Data analysis

We carried out the statistical analysis with SPSS version 27. There were no missing data. Univariate statistics were run to examine the means and standard deviations of each item as well as to check for possible outliers or entry errors. No outliers or entry errors were found. For the sake of brevity and following APA norms, detailed statistics presented in tables are not repeated in the text. Scale acronyms and item codes will be used for the result description.

5.4. Results

5.4.1. Factor analysis and reliability

We ran an Exploratory Factor Analysis (EFA) (principal axis factoring) on the 16 items composing the experimental version of the scale with a direct oblimin rotation (which allows the resultant factors to be correlated). The suitability of EFA was assessed prior to analysis. Our sample size of $n = 406$ for this analysis provided a ratio of 25 cases per item, thus largely fulfilling the recommended person-to-item ratio of 5:1 or 200 recommended cases (Howard, 2016). The overall Kaiser-Meyer-Olkin (KMO) measure was 0.95, which is 'marvelous' according to Maibom (2017)'s classification. Bartlett's Test of Sphericity was statistically significant $\chi^2(120) = 3559.288$, $P = 0.000$, indicating that the data were likely

factorizable. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.30. The communalities (proportion of item's variance explained by the extracted factors) were all above 0.30, confirming that each item shared some common variance with other items.

The EFA provided two factors based on components with Eigenvalues greater than 1. A visual exploration of the scree plot combined with a parallel analysis (O'Connor, 2000), however, suggested a three-factor retention choice, explaining 63% of the variance. Contrary to our assumptions, the data did not reflect four dimensions. We carefully examined each item and applied a combination of criteria for factor loading cutoff (as suggested by Howard (2016)). We considered satisfactory variables that loaded onto their primary factor above 0.40 and loaded onto alternative factors below 0.30. We also considered differences between the primary and alternative factor loadings. We examined the pattern of item loadings within each factor to interpret their meaning and based our factor interpretation on empirical relationships between items. Section 5.5 discusses the discrepancies between our empirically derived factors and the theoretically expected dimensions. The final selection was composed of nine items loading on the three components *Emotional interest* (EI), *Personal experience* (PE) and *Self-awareness* (SA). We found a meritorious KMO (Maibom, 2017) of 0.88 and a statistically significant Bartlett's Test of Sphericity, $\chi^2(36) = 1653.949$, $P = 0.000$. Component loadings and communalities of the rotated solution are presented in Table 2. This three-factor solution explains 72% of the variance in the data. The factor pattern loadings ranged from 0.44 to 0.86.

Estimates of internal consistency for the EMPA-D Total and each of the factors were measured by the McDonald's Coefficient Omega (Table 3). Omegas of 0.89, 0.76 and 0.72 were obtained for the EMPA-D Total, EI, PE and SA. These initial estimates of reliability suggest that the EMPA-D and its three factors have good or acceptable levels of internal consistency in the current sample.

5.4.2. Factor intercorrelations

Relationships were examined between all components within the EMPA-D. The intercorrelations among the three factors presented in Table 4 suggest that the factors are interrelated to a moderate to high level.

5.4.3. Relation with sociodemographic variables

We examined the relations between demographic variables and the EMPA-D scores. There was no significant correlation between EMPA-D and seniority in the company, $r = -0.07$, $P = 0.15$ or EMPA-D and age, $r = -0.087$, $P = 0.079$. Aligned with prior literature (Baron-Cohen and Wheelwright, 2004; Jolliffe and Farrington, 2006), an independent samples t-test revealed gender differences in the empathy scores, with women ($M = 5.30$, $SD = 0.98$) scoring higher than men ($M = 5.09$, $SD = 1.04$), $t(403) = -2.07$, $P = <0.039$, $\eta^2 = -0.21$. We conducted one-way between-subject ANOVAs to compare the respondents' profiles on the level of empathy. The empathy scores are higher in customer-facing profiles ($M = 5.31$, $SD = 0.96$) than in general service staff ($M = 5.24$, $SD = 1.03$), $F(2, 403) = 7.74$, $P = <.001$, $\eta^2 = 0.04$ (inter-group differences tested by *post hoc* analysis significant at $P = <0.001$ level). These differences provide first evidence of the scale sensitivity, by showing the scale's capacity to discriminate between groups that have known differences on the empathy construct.

5.5. Discussion

Study 1's findings show that our scale covered three dimensions out of the four initially identified in design research literature,

Table 2. Extracted components and factor loadings based on principal axis factoring with Oblimin rotation ($n = 406$)

Initial item code	Factors			Item-total correlation	Communalities	Mean	SD
	1	2	3				
<i>Emotional interest (EI)</i>							
EI1	0.86			0.82	0.81	5.33	1.53
EI3	0.80			0.77	0.68	5.33	1.49
EI4	0.85			0.78	0.73	5.16	1.5
<i>Personal experience (PE)</i>							
PE2		0.65		0.62	0.59	4.60	1.45
PE3		0.87		0.63	0.67	4.98	1.53
PE4		0.44		0.52	0.37	4.54	1.6
<i>Self-awareness (SA)</i>							
SA2			0.66	0.49	0.39	5.73	1.28
SA3			0.52	0.52	0.45	5.45	1.25
SA4			0.74	0.58	0.59	5.55	1.3
Eigenvalue	4.52	1.10	0.85				
% variance explained	50.24	12.25	9.46				

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Component loadings <0.40 are suppressed. Major loadings for each item are bolded.

Table 3. Descriptive Statistics and Omega Reliabilities of the three subscales (EI, PE, SA) and EMPA-D total scores

Scale	Nb items	Min	Max	M	SD	ω
EMPA-D Emotional interest	3	1	7	5.27	1.37	0.89
EMPA-D Personal experience	3	1	7	4.71	1.25	0.76
EMPA-D Self-awareness	3	1.33	7	5.58	1.02	0.72
EMPA-D Total	9	1.44	7	5.18	1.02	0.87

Table 4. Intercorrelations between subscales and total scores of EMPA-D

Scale	1	2	3	4
1. EMPA-D Emotional interest	-	0.61**	0.56**	0.89**
20. EMPA-D Personal experience		-	0.46**	0.84**
30. EMPA-D Self-awareness			-	0.77**
4. EMPA-D Total				-

** . Correlation is significant at the 0.01 level (two-tailed), * . Correlation is significant at the 0.05 level (two-tailed), $P < 0.01$

namely (from strongest to weakest in the amount of variance the factors contribute to): the *Emotional interest* (3 items), the *Personal experience* (3 items) and the *Self-awareness* (3 items) (see EMPA-D 9-item version in Appendix B). The EI dimension appears as more prominent in percentage of variance explained. The *Sensitivity/Immersion* in the users' world described in design frameworks (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b; Smeenk et al., 2019, 2018), is not covered by the resulting items. Qualitative analysis showed that items in this dimension were rather jargoneous (e.g. 'sensitive', 'point of reference', 'immerse', 'go to the field') or referred to specificities of the job missions (S4, S5) that were not always under the control of respondents. Despite reformulation efforts at the expert review and pretest stages, the empirical evaluation of items identified weaknesses related to jargon that resonate with prior observations.

In addition, statistical analyses showed that two items from S subscale possibly loaded on EI. Described in design literature as the second stage of the empathic process (Kouprie and Sleeswijk Visser, 2009; Smeenk et al., 2019), we cannot conclude whether the items as formulated in the S dimension were less adequate for measuring employees' empathy (who have fewer opportunities to immerse themselves in the field) than designers' empathy or whether the item construction was flawed. Initially, the items

were formulated to measure the immersion in the users' world, but the literature also simply defines this dimension as the act of taking the users' perspective (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b). The ability to take others' perspective beyond the initial emotional interest stage is a key factor of empathy (Smeenk et al., 2016; Hess and Fila, 2016b; Chang-Arana et al., 2022). We thus iterate a new set of items on this dimension in Study 2.

6. STUDY 2: NEW ITERATION OF THE SCALE AND VALIDITY TESTING

The *Sensitivity/Immersion* content area described in design empathy frameworks (Kouprie and Sleeswijk Visser, 2009; Smeenk et al., 2019) was not successfully measured after Study 1. The items were too jargoneous or non-applicable, with two of them loading possibly on the *Emotional Interest* dimension. The items were originally formulated to measure the immersion in the users' world, but the literature more globally defines this dimension as the act of taking the users' perspective (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b). As the perspective-taking component plays an essential role (Surma-aho et al., 2019; Chang-Arana et al., 2022) and often relates to the *Emotional Interest* dimension, we

created a new pool of items and conducted a second-scale validation study. This study also includes measures of convergent and discriminant validity.

6.1. Creation and pretest of items on perspective-taking

We reviewed empathy scales including a *Perspective-taking* dimension in psychology (among others; Interpersonal Reactivity Index (IRI) (Davis, 1980), Jefferson scale (Hojat et al., 2001), Empathy Assessment Index (Lietz et al., 2011)), as well as frameworks and scales in service and design literature (Wieseke et al., 2012; Smeenk et al., 2016; Surma-aho et al., 2019). Two authors individually wrote new items (28 in total). Discussing, improving and removing some of them led to a purposively large pool of 24 items about perspective-taking, including both what Surma-aho et al. (2019) define as evidence-based perspective-taking ('understanding based on primary research and user interaction') and anticipatory perspective-taking ('comprises more interpretative processes, such as attempts to foresee how users would react to new products'). Two expert reviewers from Study 1 assessed the face validity of the 24 items. They also evaluated the understanding of items by commenting on them orally. After this expert evaluation, 15 items related to *Perspective-taking* remained. We pre-tested these initial 15 items with $n = 8$ participants (2 men, 5 women, 1 undisclosed, aged from 28 to 46 ($M = 32$), fluent English speakers) through online cognitive interviews. We recruited participants working in various service domains. These were different than the participants conducting the pre-test in Study 1. Same as for Study 1, the participants assessed each item's understandability on a 5-point scale. The 15 items all reached high understandability scores above $M = 4$. The items (PT1, PT4, PT5, PT6, PT7) with the lowest scores and most discussed by the participants, were examined closely. The word 'usually' was removed from all items, as it tended to bias the respondents by mitigating the commitment of the statement. We made minor adjustments to most items based on the respondents' comments, resulting in 15 revised items (Appendix C).

6.2. Administration

We administered the 15 new items of *Perspective-taking* along with EMPA-D 9 items retained in Study 1. In addition to this dimension revision, Study 2 aimed at furthering the examination of the psychometric properties of this revised EMPA-D (construct, convergent and discriminant validity). Potential links with additional concepts such as relationships with users, user-centered practices or motivation to take action for the users (Lallemand et al., 2022) were also explored.

6.2.1. Participants

Study 2 involved $n = 305$ respondents (155 men, 149 women, 1 undisclosed). Thirteen respondents were excluded due to failing the attention check items. The age ranged from 19 to 68 ($M = 39.83$, $SD = 12.11$). 69% of the respondents held a university degree. As in Study 1, we recruited full-time service employees through the crowdsourcing platform Prolific.co. The time spent in their current organization ranged from less than a year to 39 years ($M = 8.69$, $SD = 7.74$). Participants gave informed consent and received fair compensation. All respondents indicated on a scale from 1 to 7 to what extent they work in direct contact with users (Min = 1, Max = 7, $M = 5.83$, $SD = 1.70$). Participants also rated their relationship with users (on a 7-point Likert-scale: 1—'Strongly disagree', 7—'Strongly agree') and the user-centered practices of their company (Table 5).

6.2.2. Material, procedure and data analysis

The procedure is presented in Fig. 2. Respondents were asked to fill in the scale including the new *Perspective-taking* items. Additional scales were used to assess the convergent and discriminant reliability. To avoid introducing bias in the measurement, these scales were added after the EMPA-D items, and respondents were not able to navigate back to their previous answers. Regarding convergent validity, we administered a single-item measure of empathy (SITES) (Konrath et al., 2018), the three subscales of the Interpersonal Reactivity Index; *Perspective-taking*, *Empathy concern* and *Fantasy scale* (Davis, 1980) and the Empathy Quotient (Baron-Cohen and Wheelwright, 2004). We predicted a significant low to moderate correlation between EMPA-D and these instruments. For the discriminant validity, the literature establishes that narcissism is associated with reduced empathy (Hepper et al., 2014; Burgmer et al., 2021). We administered the narcissist single-item (SINS) (Konrath et al., 2014) and the Narcissistic Personal Inventory (NPI-13) (Gentile et al., 2013). We also administered the Balanced Inventory of Desirable Responding (BIDR-16) (Hart et al., 2015). We predicted negative correlations between EMPA-D and these instruments.

Single Item Trait Empathy Scale (SITES) (Konrath et al., 2018). We administered the single-item SITES to support the construct validity of EMPA-D. On a 7-point Likert-scale (1—'Not very true of me', 7—'Very true of me'), the participants rated the following statement: 'I am an empathetic person (Note: An empathetic person understands others' feelings, and experiences care and concern for them)'.

Interpersonal Reactivity Index (IRI) (Davis, 1980). We used all the dimensions of IRI (Cronbach's $\alpha = 0.85$); *Perspective-taking* (PT—tendency to adopt the psychological point of view of others), *Empathic concern* (EC—feelings of sympathy and concern for others), *Personal distress* (PD—'self-oriented' feelings of personal anxiety and unease in tense interpersonal settings) and *Fantasy scale* (FS—tendencies to transpose themselves imaginatively into the feelings and actions of fictitious characters). Each subscale entails seven statements to which respondents express their degree of agreement on a 5-point Likert scale ranging from 1—'Does not describe me well' to 5—'Does describe me well.' After a reliability analysis, items of each subscale were averaged to create a PT score $\alpha = 0.81$, EC $\alpha = 0.82$, PD $\alpha = 0.85$ and FS $\alpha = 0.80$ following the scoring principles described in Davis (1980).

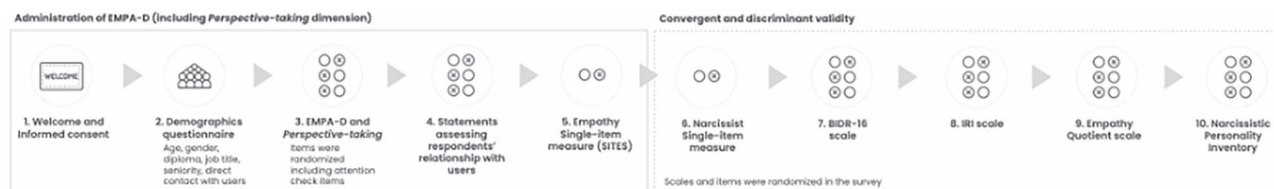
Empathy Quotient (EQ) (Baron-Cohen and Wheelwright 2004). This self-administered questionnaire measures empathy in adults. Respondents shared their agreement, on a 4-point Likert-scale from 1—'Strongly agree' to 4—'Strongly disagree', with 40 statements featuring different situations. After a reliability analysis (Cronbach's $\alpha = 0.90$), we computed an empathy quotient (Baron-Cohen and Wheelwright, 2004).

Single Item Narcissism Scale (SINS) (Konrath et al., 2014). SINS invites participants to rate, on a 7-point Likert-scale (1—'Not very true of me', 7—'Very true of me'), the following statement: 'I am a narcissist (Note: The word "narcissist" means egotistical, self-focused, and vain)'.

Narcissistic Personal Inventory (NPI-13) (Gentile et al., 2013). This widely used scale of trait narcissism includes three subscales: *Leadership/Authority* (LA, Cronbach's $\alpha = 0.65$), *Grandiose/Exhibitionism* (GE, Cronbach's $\alpha = 0.64$), *Entitlement/Exploitativeness* (EE, Cronbach's $\alpha = 0.51$). Despite its relatively low level of internal consistency documented in prior work (Ackerman et al., 2011), the sub-dimension of narcissistic entitlement (NPI EE) is particularly relevant as a divergent measure for empathy. We used the short 13-item version of NPI to limit participants' fatigue. For each

Table 5. Descriptive statistics of the statements assessing respondents' relationship with users

Statement	Min	Max	M	SD
I have good knowledge of how users experience the service	2	7	5.74	0.95
I am able to help improve the experience of service users	1	7	5.68	1.22
My job contributes to or affects the experience of the users	1	7	6.03	1.07
I am, myself, a user of the product or the service provided by my company	1	7	4.21	2.14
I am often attending trainings related to user	1	7	4.68	1.72
My company shares with me (some form of) data about the users	1	7	5.53	1.43
I am willing to co-construct the service with the users	1	7	5.23	1.29
I am ready to change my work routines and procedures to improve users' experiences	1	7	5.27	1.32
I feel responsible for the experience of the users with the service	1	7	5.53	1.29

**Figure 2.** A synthetic view of the scale administration procedure in Study 2

item, respondents chose which of two statements they most agreed with. After a reliability analysis (Cronbach's $\alpha = 0.77$), we computed a global narcissism score.

Balanced Inventory of Desirable Responding-16 (Paulhus, 1991; Hart et al., 2015). The BIDR-16 was used as a measure to control for response bias and for its theoretical links to the concept of empathy. It includes two subscales: *Self-Deceptive enhancement* (Cronbach's $\alpha = 0.78$)—i.e. honest but overly positive responding and *Impression management* (Cronbach's $\alpha = 0.78$)—i.e. bias toward pleasing others. Our participants filled out the BIDR-16 items, on a 7-point Likert scale from 1—'Strongly disagree' to 7—'Strongly agree'. After a reliability analysis (Cronbach's $\alpha = 0.83$), items were averaged into a socially desirable responding score.

6.2.3. Data analysis

The data analysis was carried out in the same way as Study 1 (see Section 5.3). There were no missing data. Univariate statistics were run to examine the means and standard deviations of each item as well as to check for possible outliers or entry errors. No outliers or entry errors were found.

To estimate the fit of the data to the scale, a confirmatory factor analysis (CFA) with maximum likelihood estimation was carried out using the *lavaan* R package. Several indices were consulted: the CFI (comparative fit index) and TLI (Tucker–Lewis index) for which a score above 0.90 indicates an acceptable fit; the SRMR (Standardized Root Mean Square Residual) and RMSEA (Root Mean Square Error of Approximation) for which a score below 0.08 indicates an acceptable fit (Jöreskog and Sörbom, 1993; Byrne, 2010; Kline, 2011).

6.3. Results

6.3.1. Factor analysis and reliability

We ran an Exploratory Factor Analysis (EFA) (principal axis factoring) on the 24 items included in Study 2 with a direct oblimin rotation (which allows the resultant factors to be correlated). Our sample size ($n = 305$) largely fulfills the recommendation of a minimum sample size of 200 or 5-to-1 participant-to-variable ratio (Howard, 2016). We found a marvelous KMO (Maibom, 2017) of 0.97 and a statistically significant ($P = 0.000$) Bartlett's Test

of Sphericity, $\chi^2(276) = 5321.294$, $P < 0.001$. All variables had at least one correlation coefficient greater than 0.30 in the correlation matrix. The communalities (proportion of item's variance explained by the extracted factors) were all above 0.30, confirming that each item shared some common variance with other items.

The EFA initially provided three possible factors (based on components with Eigenvalues > 1), explaining 64% of the variance in the data. A visual exploration of the scree plot combined with a parallel analysis (O'Connor, 2000) confirmed this factor retention choice. Contrary to our assumptions, the data did not reflect a potential distinction between *Emotional interest* (EI) and *Perspective-taking* (PT).

We carefully examined each item and applied a combination of criteria for factor loading cutoff (as suggested by Howard (2016)). We considered satisfactory variables that loaded onto their primary factor above 0.40 and loaded onto alternative factors below 0.30. We also considered differences between the primary and alternative factor loadings. Additional arguments for deleting these items resulted from pretest comments (see Section 6.1). The final three-factor solution (explaining 74% of the variance in the data) is composed of 11 items loading on the three components *Emotional interest/Perspective-taking*, *Personal experience* and *Self-awareness*. We found a marvelous KMO (Maibom, 2017) of 0.91 and a statistically significant Bartlett's Test of Sphericity, $\chi^2(55) = 2036.890$, $P = 0.000$. Out of the new items introduced, three were retained (PT3, PT5, PT9) and completed the EI dimension to become EIPT. Table 6 presents the component loadings and communities of the rotated solution. The factor pattern loadings ranged from 0.43 to 0.98. Reliability analysis for EMPA-D three subscales is reported in Table 8 (using McDonald's Coefficient Omega, except for PE, which has fewer than three items) and suggests good levels of internal consistency in the current sample.

We ran a CFA based on the EFA results, including the three components and their respective items. The results indicated an adequate fit of the data to the model: $\chi^2(41) = 127.636$, $P < 0.001$, CFI = 0.957, TLI = 0.942, RMSEA = 0.083, SRMR = 0.046. The coefficients between items and components are all statistically significant and are reported in Table 7.

Table 6. Extracted components and factor loadings based on principal axis factoring with Oblimin rotation ($n = 305$)

Initial item code	Rotated component			Item-total correlation	Communalities	Mean	SD
	1	2	3				
<i>Emotional interest/Perspective-taking (EIPT)</i>							
EI1	0.98			0.87	0.84	5.56	1.31
EI3	0.84			0.79	0.69	5.58	1.33
EI4	0.88			0.82	0.74	5.48	1.41
PT3	0.82			0.80	0.69	5.76	1.38
PT5	0.68			0.78	0.66	5.41	1.34
PT9	0.66			0.76	0.64	5.64	1.35
<i>Personal experience (PE)</i>							
PE2		0.78		0.61	0.62	4.76	1.42
PE3		0.78		0.61	0.59	5.05	1.47
<i>Self-awareness (SA)</i>							
SA2			0.78	0.54	0.53	5.90	1.13
SA3			0.43	0.50	0.42	5.72	1.15
SA4			0.67	0.58	0.55	5.75	1.10
EigenValue	5.87	1.25	1				
% variance explained	53.33	11.39	9.07				

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Component loadings $< .40$ are suppressed. Major loadings for each item are bolded.

Table 7. Results of the CFA: coefficients, standard errors and standardized coefficients

	Coef.	Std.Err	Std. coef.
<i>Emotional interest/Perspective-taking (EIPT)</i>			
EI1	1.191	0.058	0.911
EI3	1.106	0.063	0.833
EI4	1.213	0.065	0.863
PT3	1.13	0.065	0.821
PT5	1.077	0.064	0.803
PT9	1.065	0.065	0.789
<i>Personal experience (PE)</i>			
PE2	1.157	0.084	0.817
PE3	1.09	0.086	0.743
<i>Self-awareness (SA)</i>			
SA2	0.69	0.067	0.612
SA3	0.789	0.067	0.686
SA4	0.807	0.064	0.734

Table 8. Descriptive statistics and omega reliabilities of the three subscales (EIPT, PE, SA) and total scores of the second iteration of EMPA-D

Scale	Nb items	Min	Max	M	SD	ω
EMPA-D EIPT	6	1.33	7	5.59	1.17	0.93
EMPA-D PE	2	1	7	4.90	1.29	0.76
EMPA-D SA	3	2.33	7	5.79	0.90	0.72
EMPA-D Total	11	2.78	7	5.43	0.91	0.91

Cronbach's alpha was computed for PE because the McDonald's Omega cannot be estimated with less than three items

6.3.2. Factors intercorrelations

Relationships were examined between all components within this second iteration of EMPA-D. The intercorrelations among the three factors presented in Table 9 suggest that the factors are interrelated at a moderate to high level.

6.3.3. Convergent and discriminant validity

We assessed the construct validity by investigating the reliability of subscales, scales and related constructs and the relationships between them (Table 10).

We found significant positive associations between EMPA-D Total and SITES $r = 0.42$, $P = < 0.001$, $CI(.33, 0.51)$ and between EMPA-D Total and Empathy Quotient $r = 0.34$, $P = < 0.001$, $CI(.24, .44)$, supporting the convergent validity of EMPA-D scale as a distinct measure of empathy. As predicted, associations are also positive between IRI and EMPA-D Total $r = 0.33$, $P = < 0.001$, and for the three EMPA-D subscales, with a higher correlation between IRI PT and EMPA-D EIPT $r = 0.41$, $P = < 0.001$, 17.1% of variance explained.

We predicted significant negative correlations between EMPA-D Total and the measure of narcissism SINS and NPI-13 (at the

Table 9. Intercorrelations between subscales and total scores of EMPA-D (with EIPT, PE and SA dimensions) and validity instruments

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-	0.55**	0.53**	0.86**	0.41**	-0.13*	0.34**	0.41**	0.15**	-0.07	0.30**	0.30**	0.04	-0.17**	-0.04	-0.07	0.00	0.20**	0.13*
2		-	0.36**	0.83**	0.32**	-0.11	0.26**	0.33**	0.11	-0.07	0.22**	0.23**	0.04	-0.12*	-0.08	-0.06	0.06	0.16**	0.13*
3			-	0.73**	0.29**	-0.13*	0.32**	0.37**	0.20**	-0.12*	0.28**	0.33**	0.01	-0.11	0.01	-0.03	0.09	0.17**	0.16**
4				-	0.42**	-0.15*	0.37**	0.46**	0.18**	-0.10	0.33**	0.34**	0.04	-0.16**	-0.05	-0.07	0.06	0.22**	0.17**
5					-	-0.41**	0.58**	0.46**	0.32**	-0.04	0.49**	0.56**	-0.10	-0.34**	-0.04	-0.19**	0.07	0.28**	0.21**
6						-	-0.32**	-0.27**	-0.17**	0.04	-0.26**	-0.33**	0.24**	0.37**	0.31**	0.39**	-0.02	-0.29**	-0.19**
7							-	0.52**	0.45**	0.06	0.76**	0.64**	-0.22**	-0.38**	-0.15**	-0.31*	-0.08	0.30**	0.14*
8								-	0.28**	-0.13*	0.60**	0.49**	-0.25**	-0.38**	-0.09	-0.29**	0.09	0.44**	0.32**
9									-	0.20**	0.76**	0.33**	-0.09	-0.10	-0.02	-0.09	-0.22**	-0.09	-0.18**
10										-	0.47**	-0.15**	-0.25**	0.01	-0.06	-0.14*	-0.54**	-0.25**	-0.46**
11											-	0.49**	-0.30**	-0.31**	-0.12*	-0.31**	-0.31**	0.12*	-0.10
12												-	-0.07	-0.29**	-0.09	-0.18**	0.25**	0.37**	0.37**
13													-	0.46**	-0.47**	0.83**	0.28**	-0.17**	0.05
14														-	0.32**	0.72**	-0.06	-0.39**	-0.28**
15															-	0.79**	0.23**	-18**	0.02
16																-	0.21**	-0.30**	-0.07
17																	-	0.42**	0.83**
18																		-	0.86**
19																			-

**Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level (two-tailed). 1. EMPA-D EIPT, 2. EMPA-D PE, 3. EMPA-D SA, 4. EMPA-D Total, 5. SITES, 6. SINS, 7. IRI EC, 8. IRI PT, 9. IRI FS, 10. IRI PD, 11. IRI Total, 12. EQ-40, 13. NPI-13 LA, 14. NPI-13 EE, 15. NPI-13 LA, 16. NPI-13 Total, 17. BIDR-16 SDE, 18. BIDR-16 IM, 19. BIDR-16 Total.

Table 10. Descriptive statistics and alpha reliabilities of the subscales and total scores of validity instruments

Scale	Nb items	Min	Max	M	SD	Cronbach's α
SITES	1	1	7	5.70	1.23	
SINS	1	1	7	1.78	1.21	
IRI Empathic concern	7	0.71	4	2.78	0.71	0.82
IRI Perspective-taking	7	0.86	4	2.61	0.64	0.81
IRI Fantasy scale	7	0.29	4	2.38	0.79	0.80
IRI Personal distress	7	0.00	4	1.62	0.78	0.85
IRI Total	28	0.61	3.71	2.35	0.47	0.85
EQ-40	40	10	71	41.69	12.96	0.90
NPI-13 Leadership/Authority	4	0.00	4	0.86	1.14	0.65
NPI-13 Grandiose exhibitionism	5	0.00	5	0.87	1.21	0.64
NPI-13 Entitlement/Exploitativeness	5	0.00	4	0.66	0.94	0.51
NPI-13 Total	13	0.00	12	2.39	2.57	0.77
BIDR-16 Self-deceptive enhancement	8	1.38	7.50	4.36	1.18	0.78
BIDR-16 Impression management	8	1	8	4.79	1.31	0.78
BIDR-16 Total	16	2	7.69	4.58	1.05	0.83

Table 11. Intercorrelations between sociodemographic variables and subscales of the second iteration of EMPA-D (with EIPT, PE and SA dimensions)

Items/EMPA-D dimensions	EIPT	PE	SA	EMPA-D Total
I have good knowledge of how users experience the service	0.45**	0.35**	0.34**	0.50**
I am able to help improve the experience of service users	0.50**	0.35**	0.21**	0.43**
My job contributes to or affects the experience of the users	0.46**	0.30**	0.40**	0.47**
I am, myself, a user of the product or the service provided by my company	0.11 (P = 0.053)	.20**	-.043 (P = 0.46)	.13 (P = 0.023)
I am often attending trainings related to the users	0.41**	0.34**	0.17**	0.39**
My company shares with me (some form of) data about the users	0.36**	0.17*	0.29**	0.33**
I am willing to co-construct the service with the users	0.59**	0.38**	0.25**	0.51**
I am ready to change my work routines and procedures to improve users' experiences	0.53**	0.35**	0.33**	0.51**
I feel responsible for the experience of the users with the service	0.61**	0.39**	0.35**	0.56**

**Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level (two-tailed).

Table 12. Definitions of the three dimensions measured by the second iteration of EMPA-D

Dimension	Definition
Emotional interest and Perspective-taking (EIPT)	Measures the employees' willingness to learn from users, including their interest and curiosity toward users. It also measures their active engagement in evidence-based perspective-taking
Personal experience (PE)	Measures the employees' ability to use their personal experience with the service (or the one of their acquaintances) to understand users' experience.
Self-awareness (SA)	Measures the employees' ability to distinguish between their experience and the users' experience (self/other distinction).

Table 13. Final version of the 11-item Empathy in Design Scale (EMPA-D)

Item code	Item	Initial item code
	<i>Emotional interest/Perspective-taking (EIPT)</i>	
EIPT-1	I am interested to learn about users' experiences and needs	EI1
EIPT-2	I am curious about users' experiences and needs	EI3
EIPT-3	I want to learn about users' experiences and opinions about the service	EI4
EIPT-4	As an employee, I try to find out what the users' needs are	PT3
EIPT-5	I take action to view things from the users' perspective	PT5
EIPT-6	I am actively listening to users' experiences to better understand their perspectives	PT9
	<i>Personal experience (PE)</i>	
PE-1	The experiences and feelings of users resonate with my own	PE2
PE-2	I understand the users' experiences because I know how it feels	PE3
	<i>Self-awareness (SA)</i>	
SA-1	I am aware that my experiences as an employee are different from the ones of users	SA2
SA-2	I realize that there are similarities and differences between my experiences and the ones of users	SA3
SA-3	I understand why users perceive things differently than I do as an employee	SA4

The 7-point of the scale are: 1/Does not describe me at all 2/Barely describes me 3/Somewhat describes me 4/Moderately describes me 5/Generally describes me 6/Mostly describes me 7/Completely describes me. Instructions: The table below includes statements related to your professional context. By 'users' we mean people/customers who use your organization's products or services. Please use the 7-point scale to indicate the degree to which these statements accurately describe you or not. Respond spontaneously: there are no right or wrong answers, only your perspective matters.

EE subscale level). There is a small negative correlation between SINS and EMPA-D Total $r = -0.15$, $P = 0.01$, $CI(-0.26, -0.04)$, and its dimensions EIPT and SA $r = -0.13$, $P = 0.02$ for both. There is no significant correlation with PE. Similarly, we see a small negative correlation between NPI-13 EE and EMPA-D Total, EIPT ($r = -0.17$, $P = 0.002$, 3% of variance explained) or PE ($r = -0.12$, $P = 0.04$, 1.4% of variance explained).

We found a negligible significant association between social desirability BIDR-16 and EMPA-D Total $r = 0.17$, $P = 0.003$, $CI(0.06, 0.28)$, and BIDR-16 IM $r = 0.22$, $P = <0.001$, 4.8% of variance explained. The same holds for associations between BIDR-16 Total and EMPA-D EIPT $r = 0.13$, $P = 0.03$, 1.6% of variance explained, and EMPA-D PE $r = 0.13$, $P = 0.02$, 1.8% of variance explained. This suggests that EMPA-D measures empathy rather than people's wishes to be perceived as empathic (Jolliffe and Farrington, 2006). There were no significant associations between BIDR-16 Self-deceptive enhancement and EMPA-D Total $r = 0.06$, $P = 0.32$, 0.3% of variance explained, and its subscales. These analyses provide evidence for the discriminant validity of EMPA-D.

6.4. Relation with sociodemographic variables

We examined the relations between demographic variables and the second iteration of EMPA-D scores. As in Study 1, there was no significant correlation between EMPA-D and seniority, $r = -0.06$, $P = 0.27$. There is, however, a significant correlation between EMPA-D and the extent to which respondents work in direct contact with the users of the service ($r = 0.31$, $P < 0.001$). As a reminder, participants rated their relationship with users and the user-centered practices of their company (see Table 5). It is interesting to investigate the relationship between these variables and the empathy dimensions (Table 11). Aligned with our assumptions,

the fact for an employee to be a user of the service is positively correlated to the PE dimension ($r = 0.20$, $P < 0.001$) but not to EIPT (0.11 , $P = 0.053$) or SA ($r = -0.043$, $P = 0.456$). As a user, the experience of other users resonates with one's own (Bove, 2019). Being a user of the service is, however, not a good predictor of the overall empathy toward users, with only 1.7% of the variance explained when conducting a linear regression, $F(1, 303) = 5.196$, $P = 0.23$, $R^2 = 0.17$.

A challenge related to empathy in design is also to make everyone in the organization realize the impact of their job on the service experience (Postma et al., 2012) even if they are not customer-facing. A linear regression was run to predict EMPA-D from the perceived ability to help improve the experience of the service. This variable statistically significantly predicted EMPA-D Total, $F(1, 303) = 95.783$, $P < 0.001$, $R^2 = 0.240$, explaining 24% of the variance in empathy level. Similarly, a multiple regression was run to predict EMPA-D total from user-centered communication or training actions performed by a company (items 'my company shares with me (some form of) data about the users' and 'I am often attending trainings related to the users'). These variables statistically significantly predicted EMPA-D total, $F(2, 302) = 38.011$, $P < 0.001$, $R^2 = 0.201$, explaining 20% of the variance in empathy level. All two variables added statistically significantly to the prediction, $P < 0.001$.

Aligned with prior work, we then researched whether empathy toward users would impact the employees' willingness to take action and contribute to the improvement of the service experience. A linear regression was run to predict the readiness for action (item 'I am ready to change my work routines and procedures to improve users' experiences') from EMPA-D total. The empathy level statistically significantly predicted readiness

for action, $F(1, 303) = 103.622$, $P < 0.001$, $R^2 = 0.255$, with 25% of the variance explained. It is beyond the scope of the present work to examine these relationships further, yet it is useful to note that other mediating factors could be explored in future work and that EMPA-D can be a useful tool in this endeavor.

6.5. Discussion

In Study 2, we adapted the *Sensitivity/Immersion* dimension to employees through new items oriented on perspective-taking. Rather than a complete pivot away from the SI dimension, these items reflect the active immersion in the perspective of users, in the way designers immerse themselves in the users' world (Kouprie and Sleeswijk Visser, 2009; Smeenk et al., 2019). Surma-aho et al. (2019) refer to evidence-based perspective-taking based on primary research and user interaction. The findings from Study 2 thus result from an improved contextualization of employees' roles and how they relate to the frameworks initially focused on designers.

The analyses conducted in Study 2 show that this second iteration of EMPA-D covers three dimensions of empathy in design: *Emotional interest/Perspective-taking*, *Personal experience* and *Self-awareness* (see Table 12 and Appendix D). The *Emotional interest/Perspective-taking* items load on a unique dimension as theorized by Hess and Fila (2016b) (and previously observed in Study 1 with overlaps between EI and S items). In practice, the *Emotional interest* drives the action of *Perspective-taking*; the concepts are thus related and inseparable. The scale eventually includes more items for this first dimension, which reflects a larger breadth of content of this dimension as compared to the other two.

Analyses of factor loadings led us to eliminate PE4 from the final solution (part of the PE dimension in Study 1). PE4 loaded more strongly with the EIPT dimension. Its formulation—'I compare users' experiences with the ones of people I know'—was likely problematic as the item implicitly assumes that the respondent personally knows someone using the service.

Regarding convergent validity, we found significant positive associations between EMPA-D Total and SITES, and with Empathy Quotient. Significant correlations between EMPA-D EIPT and both IRI PT and IRI FS illustrate the adequate measure of perspective-taking by our dedicated EMPA-D EIPT subscale. Regarding discriminant validity, our results align with prior literature (Hepper et al., 2014; Burgmer et al., 2021); EMPA-D is negatively correlated with narcissism, and more specifically with NPI-13 *Entitlement/Exploitativeness*. We found evidence in the relation between EMPA-D and BIDR-16 that EMPA-D items are not overly subject to social desirability (Jolliffe and Farrington, 2006).

7. GENERAL DISCUSSION

In this paper, we presented the development and initial validation of the EMPA-D through two studies ($n = 406$ and $n = 305$). Drawing on the vast literature and measurement scales in psychology, we propose a specific self-report tool adapted to research and practice in service design (Table 13).

The dimensions and items of EMPA-D are grounded in design frameworks initially created to investigate designers' empathy toward users (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b; Smeenk et al., 2018, 2019), which we extended in the context of service design to internal stakeholders (e.g. employees) who are instrumental to the users' service experience (Stickdorn et al., 2018). In this section, we discuss the possible scenarios for the use of the EMPA-D Scale, nuance the relevance of the

quantitative approach of empathy in design, acknowledge the limitations of the present studies and present opportunities for future work.

7.1. What are possible use scenarios for the EMPA-D scale?

EMPA-D is thought for any service context—private or public, in all sectors—addressing a variety of professionals involved in shaping the service experience. Its theoretical foundations and generic formulation make it likely suitable in the product design domain too, yet it has not been validated in this context. How can the scale be worthwhile to service design? The scale is an opportunity to increase the credibility of an empathic approach in the industry, which valorizes key performance indicators and return-on-investment to guide decision-making (Bias and Mayhew, 2005; Stickdorn et al., 2018). The company can address questions such as: on which empathic dimension do specific groups of employees score the highest or lowest? Is there a high emotional interest for the users? Is this interest shared across the organization or rather limited to customer-facing employees? What are the antecedents and outcomes of empathy in design (Surma-aho et al., 2019)? The data allow organizations to identify staff members that are highly empathic toward customers and who might be suitable to take on a more active role in the design process. Informed by these data, the organization can furthermore deploy targeted empathic design interventions (Drouet et al., 2023; Drouet et al., 2024). Contrary to the psychological construct of empathy as a personality trait, which has led to some criticisms in the community, empathy in design is better understood as a skillset, including attitudes, skills and knowledge (cf. the empathic formation concept by Smeenk et al., 2019). Methods such as co-design involving service employees and users are, for instance, known to support perspective-taking (Smeenk et al., 2016; Akoglu and Dankl, 2021). Hence, direct contact was shown by van Rijn et al. (2011) to be the most effective to prompt designers' empathy toward their users and customers. Emotional interest can be prompted through design documentaries (Raijmakers et al., 2006) or physical journey maps (Lallemand et al., 2022), which both communicate rich and multidimensional accounts of people's lived experiences.

Our findings also emphasize the importance of diffusing a culture of shared responsibility for the service experience: employees recognizing their ability to improve the service experience appears as an antecedent to developing empathy (and especially emotional interest). EMPA-D can then also be used to efficiently assess the effectiveness of the intervention (Drouet et al., 2022; Drouet and Lallemand, 2024) by measuring the empathy levels of the service employees before (baseline measure), during or after the empathic design intervention.

7.2. A nuanced approach to empathy in service design

Many factors come into play when addressing empathy in design and the overall organizational strategy and resources should be considered when using an empathy measurement tool. At the stakeholder level, the need to be empathic toward users varies with the job mission. Some dimensions of empathy might be useful to all service employees (e.g. *Emotional interest*), while others can be specific to customer-facing personnel (e.g. appealing to *Personal experience*). Empathy can also be limited by cultural and demographic factors (Chang-Arana et al., 2022). The EMPA-D scores thus need to be interpreted according to the service context and population surveyed and put into perspective with qualitative data.

Potential ethical risks related to measuring empathy in companies must also be considered (Drouet et al., 2022). First, an empathy measurement tool could be diverted from its primary use and be used to assess employees' performance. If the goal is the selection and deployment of empathic interventions, we suggest deriving analyses at the team level and safeguarding the anonymity of individuals. This also calls for interventions at the organizational level (Drouet et al., 2024), considering empathy as a skill to be developed rather than a personality concern as it is common in healthcare (e.g. Pratiwi et al., 2022) or engineering domains (Hess et al., 2017; Surma-aho et al., 2018; Walther et al., 2020; Alzayed et al., 2021). This view on empathy as a skill is aligned with our analyses in Study 2, which show that 20% of the variance in reported employees' empathy level can be explained by user-centered actions performed by a company. Empathy training initiatives like Google's program Pokerface (Liu et al., 2018) or Vasques et al. (2022)'s love and breakup letters as used by Drouet and Lallemand (2024) can be a source of inspiration. A second risk is that empathy could be perceived by companies as the holy grail of user experience, rather than improving the service through a holistic user-centered approach. Empathy scores might turn into an end in itself with a marketing objective. Some authors criticize the adverse side of empathy when designing in a capitalistic logic (Holt, 2011), while others point out that empathy in design has become more an ideology than a helpful principle (Heylighen and Dong, 2019). This can result in the empathy trap phenomena (Mattelmäki et al., 2014) where designers forget their own concerns. Following the dialogical approach to empathy by Wright & McCarthy (2008), each person engages from their own perspectives and appreciates the other's perspective as other. 'In an empathic relationship the "designer" does not relinquish his/her position to "become the user", a position from which nothing new can be created, rather the designer responds to what they see as the user's world from their own perspective as designer'. A similar dialogical approach should be initiated when expanding the development of empathy to the service employees.

Finally, in industry, our tool aims at taking stock of empathy as a parameter of the company's user-centeredness and supporting employees (at large, not only customer-facing) in building empathy toward users of a service or product. It might thus seem desirable for an employee to showcase empathic tendencies and 'cheat the test' (Hemmerdinger et al., 2007) by presenting themselves in a positive light. Measures like the BIDR-16 (Hart et al., 2015) or the Lie scale of the Eysenck Personality Scale (Eysenck and Eysenck, 1991) can be used to make self-report instruments robust to social desirability. Other best practices include anonymization (and reassuring communication) in the administration of the scale and a clear vision shared by the management that the measure is meant for self-growth rather than the assessment of personnel.

7.3. Limitations

This study involves several limitations. Regarding the development of the scale, the initial items in study 1 constituted a good representation of the conceptual facets of empathy in design as documented in the literature. Nevertheless, the item pool might appear small because we chose to rather pilot test and refine it during cognitive interviews. The first pool would also have benefited from less active formulations regarding the immersion dimension. This insight led to the development of new items on the perspective-taking dimension in Study 2. In our scale development process, some items were removed due to the lack of applicability to the roles of some service employees (as compared

to, e.g. designers, customer-facing personnel or simply depending on their job missions). While removing items for this purpose is common practice in scale development, one can challenge whether these items did not reflect the manifestation of empathy in service employees or whether this manifestation is expressed in a variety of ways insofar that it may be challenging to capture in a concise and generic instrument.

Another limitation related to our sample of paid crowd workers. If studies on these samples do not discredit their quality (Behrend et al., 2011; Peer et al., 2022), their motivation and the authenticity of their answers might be influenced by this status. We included attention check items, as a common strategy to cope with issues of inattentiveness and satisficing.

Although self-report instruments are efficient in collecting data, this approach is prone to biases (Lietz et al., 2011). Two of them are of particular attention: the acquiescence bias and the social desirability bias. A mix of positively and negatively worded items can reduce the acquiescence bias. We did not adopt this strategy, in an effort to maximize understandability: rating negative phrasing requires more attention from the respondents (Swain et al., 2008; Weijters and Baumgartner, 2012). Social desirability is another concern when the results of the test might be perceived as a measure of performance (which is mitigated here by the sample of crowdworkers).

7.4. Future work

At the construct level, researchers (Chang-Arana et al., 2022; Surma-aho and Hölttä-Otto, 2022) advocate that quantifying empathy in design would only be possible once consensus has been reached within the community on its scope and definition. These processes are cumulative and slow and require scientific consolidation efforts. The EMPA-D is a first step, paving the way to further initiatives and validation studies. Empirical use of the scale *in situ* through various design interventions will contribute new insights into the usefulness and applicability of the tool and allow it to improve iteratively. It is also necessary to conduct follow-up studies to gather additional validity evidence and further refine the scale's psychometric properties and its robustness. Future work on scale sensitivity could analyze variations in empathetic responses to different scenarios. Studies on the predictive validity of EMPA-D should be conducted next to investigate whether the scale helps to better predict the service experience or other service-related outcomes. We invite other researchers to pursue work of quantifying empathy in design, e.g. by using EMPA-D in other contexts or as a complement of other measures. Some scales in the healthcare domain exist in two parallel versions (Davis, 1980; Hojat et al., 2001): one for employees and one for users to rate their perception of the employees' empathy toward them. As empathy occurs in interaction and it 'takes two to tango' (Ngo et al., 2020), it is equally relevant to investigate the role of users' empathy toward employees on the resulting service experience (Wieseke et al., 2012; Ngo et al., 2020). Beyond the measurement of empathy, it is worthwhile in future work to study potential interventions related to the development of empathic understanding within companies (Drouet et al., 2023; Drouet et al., 2024), such as empathic training and to see which empathic design methods can be deployed with service employees (Drouet et al., 2024; Drouet and Lallemand, 2024).

8. CONCLUSION

In this paper, we report on the development and first validation of the EMPA-D, a short 11-item measure of empathy in service

design. The scale produces a total empathy score, and scores for three dimensions, namely, *Emotional interest/Perspective-taking*, *Personal experience* and *Self-awareness*. The creation of this tool contributes to filling the gap in metrics to assess empathy in the service design context, by operationalizing the dimensions of empathy as described in design research frameworks (Kouprie and Sleeswijk Visser, 2009; Hess and Fila, 2016b; Smeenk et al., 2018, 2019). It can support the collection of empirical evidence through experimental research designs (Surma-aho and Hölttä-Otto, 2022) that will allow refining or consolidating theoretical claims. In industry, the scale intends to support organization to assess employees' empathy toward users and consequently deploying empathic methods to foster user-centeredness.

ACKNOWLEDGEMENTS

We would like to thank the HCI research team at the University of Luxembourg, and particularly Sophie Doublet, Verena Distler and Margault Sacré, who contributed to the scale validity analyses. We also thank Vincent Koenig, who contributed to the funding acquisition and initial supervision of the dissertation project related to this work, as well as Sophie Lacour, Tom Nickels and the Luxembourgish Railway Company CFL for supporting this project.

AUTHOR CONTRIBUTIONS

Conceptualization, L.D. and C.L.; Methodology, L.D. and C.L.; Validation, L.D., C.L., K.B.; Formal analysis, L.D., C.L., K.B.; Investigation, L.D. and C.L.; Writing—Original Draft, L.D., C.L., K.B.; Writing—Review & Editing, L.D., C.L., K.B.; Project administration, L.D. and C.L.; Supervision, C.L.; Funding acquisition, C.L., Vincent Koenig (acknowledged). All authors have read and agreed to the published version of the manuscript.

FUNDING

This study is supported by the Luxembourgish Railway—CFL, as part of a research project whose objective is to increase company stakeholders' empathy toward passengers. The funding source has no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author.

References

- Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W. and Kashy, D. A. (2011) What does the Narcissistic Personality Inventory really measure? *Assessment*, **18**, 67–87. <https://doi.org/10.1177/1073191110382845>.
- Akoglu, C. and Dankl, K. (2021) Co-creation for empathy and mutual learning: a framework for Design in Health and Social Care. *CoDesign*, **17**, 296–312. <https://doi.org/10.1080/15710882.2019.1633358>.
- Alzayed, M. A., McComb, C., Menold, J., Huff, J. and Miller, S. R. (2021) Are you feeling me? An exploration of empathy development in engineering design education. *J. Mech. Des.*, **143**, 112–301. <https://doi.org/10.1115/1.4048624>.
- Bahadur, W., Khan, A. N., Ali, A. and Usman, M. (2020) Investigating the effect of employee empathy on service loyalty: the mediating role of trust in and satisfaction with a service employee. *J. Relatsh. Mark.*, **19**, 229–252. <https://doi.org/10.1080/15332667.2019.1688598>.
- Baron-Cohen, S. (2004) *The Essential Difference: Male and Female Brains and the Truth about Autism*. Basic Books, New York.
- Baron-Cohen, S. and Wheelwright, S. (2004) The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and Normal sex differences. *J. Autism Dev. Disord.*, **34**, 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>.
- Battaly, H. D. (2011) “Is empathy a virtue?” In Coplan, A. and Goldie, P. (eds), *Empathy: Philosophical and Psychological Perspectives*, pp. 277–301. Oxford University Press, New York.
- Behrend, T. S., Sharek, D. J., Meade, A. W. and Wiebe, E. N. (2011) The viability of crowdsourcing for survey research. *Behav. Res. Methods*, **43**, 800–813. <https://doi.org/10.3758/s13428-011-0081-0>.
- Bias, R. G. and Mayhew, D. J. (2005) Cost-Justifying Usability An Update for an Internet Age. In *Interactive technologies edition*. Morgan Kaufmann, San Francisco.
- Bove, L. L. (2019) Empathy for service: benefits, unintended consequences, and future research agenda. *J. Serv. Mark.*, **33**, 31–43. <https://doi.org/10.1108/JSM-10-2018-0289>.
- Burgmer, P., Weiss, A. and Ohmann, K. (2021) I don't feel ya: how narcissism shapes empathy. *Self Identity*, **20**, 199–215. <https://doi.org/10.1080/15298868.2019.1645730>.
- Buttle, F. (1996) SERVQUAL: review, critique, research agenda. *Eur. J. Mark.*, **30**, 8–32. <https://doi.org/10.1108/03090569610105762>.
- Byrne, B. M. (2010) *Structural equation modeling with AMOS: Basic Concepts, applications, and programming*. Routledge, New York.
- Cabrera-Nguyen, P. (2010) Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *J. Soc. Soc. Work Res.*, **1**, 99–103. <https://doi.org/10.5243/jsswr.2010.8>.
- Castillo-Díaz, M. and Padilla, J. L. (2013) How cognitive interviewing can provide validity evidence of the response processes to scale items. *Soc. Indic. Res.*, **114**, 963–975. <https://doi.org/10.1007/s11205-012-0184-8>.
- Chang-Arana, S.-A. A., Hölttä-Otto, K. and Sams, M. (2022) Under the umbrella: components of empathy in psychology and design. *Design Science*, **8**, e20, 1–29. <https://doi.org/10.1017/dsj.2022.13>.
- Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., and Hölttä-Otto, K. (2020) Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, **6**, e16, 1–34. <https://doi.org/10.1017/dsj.2020.14>.
- Chapman, L. and Plewes, S. (2014) A UX Maturity Model: Effective Introduction of UX into Organizations. In Marcus, A. (ed), *Design, User Experience, and Usability. User Experience Design Practice*, pp. 12–22. Springer, Cham.
- Cuff, B. M., Brown, S. J., Taylor, L. and Howat, D. J. (2016) Empathy: a review of the concept. *Emot. Rev.*, **8**, 144–153. <https://doi.org/10.1177/1754073914558466>.
- Davis, M. H. (1980) A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, **10**, 85.
- Devecchi, A. and Guerrini, L. (2017) Empathy and design. A new perspective. *Des. J.*, **20**, S4357–S4364. <https://doi.org/10.1080/14606925.2017.1352932>.
- DeVellis, R. F. (2003, 4th edition 2017) *Scale Development: Theory and Applications*. SAGE, Thousand Oaks.

- Dong, Y., Dong, H. and Yuan, S. (2018) Empathy in Design: A Historical and Cross-Disciplinary Perspective. In Baldwin, C. (ed), *Advances in Neuroergonomics and Cognitive Engineering*, pp. 295–304. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-60642-2_28.
- Drouet, L. and Lallemand, C. (2024) Empathy in Service Design: Prompting Employees' Empathy with Users through Love and Breakup Declarations. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*, June 25–27, 2024, Newcastle upon Tyne, United Kingdom, p. 14. ACM, New York, NY, USA. <https://doi.org/10.1145/3663384.3663388>.
- Drouet, L., Bongard-Blanchy, K., Koenig, V. and Lallemand, C. (2022) Empathy in Design Scale: Development and Initial Insights. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 373, 1–7. <https://doi.org/10.1145/3491101.3519848>.
- Drouet, L., Sleeswijk Visser, F. and Lallemand, C. (2023) Using Empathy-Centric Design in Industry: Reflections from the UX Researcher, the Client, and the Method Expert. In *Proceedings of the 2nd Empathy-Centric Design Workshop (EMPATHICH '23)*, pp. 1–9. Association for Computing Machinery, New York, NY, USA, Article 10.
- Drouet, L., Sleeswijk Visser, F., Pagán, B. and Lallemand, C. (2024) Towards a mapping of empathic design methods. In Hekkert, P., Ciuccarelli, P., Gray, C. (eds), *Proceedings of DRS2024: Resistance, Recovery, Reflection, Reimagination*, Boston: Design Research Society.
- Eysenck, H. J. and Eysenck, S. B. G. (1991) *Manual of the Eysenck Personality Scales (EPS Adult), Comprising the EPQ-revised (EPQ-R) (including Addiction and Criminality Scales) EPQ-R Short Scale Impulsiveness (IVE) Questionnaire (Impulsiveness, Venturesomeness, Empathy)*. Hodder and Stoughton, London. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '04)*, pp. 261–268. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/1013115.1013152>.
- Fulton, Suri, J. (2003) *Empathic Design: Informed and Inspired by Other People's Experience*, In Koskinen, I., Battarbee, K. and Mattelmäki, T. (eds), *Empathic Design: User Experience in Product Design*, pp. 51–65. IT Press, Finland.
- Furr, R. M. and Bacharach, V. R. (2013) *Psychometrics: An Introduction* (2nd). SAGE Publications, London, UK.
- Gentile, B., Miller, J. D., Hoffman, B. J., Reidy, D. E., Zeichner, A. and Campbell, W. K. (2013) A test of two brief measures of grandiose narcissism: the narcissistic personality inventory–13 and the narcissistic personality inventory–16. *Psychol. Assess.*, **25**, 1120–1136. <https://doi.org/10.1037/a0033192>.
- Hart, C., Ritchie, T., Hepper, E. and Gebauer, J. (2015) The balanced inventory of desirable responding short form (BIDR-16). *SAGE Open*, **5**, 215824401562111. <https://doi.org/10.1177/2158244015621113>.
- Hemmerdinger, J. M., Stoddart, S. D. and Lilford, R. J. (2007) A systematic review of tests of empathy in medicine. *BMC Medical Education*, **7**, 24. <https://doi.org/10.1186/1472-6920-7-24>.
- Hepper, E. G., Hart, C. M. and Sedikides, C. (2014) Moving narcissus: can narcissists be empathic? *Personal. Soc. Psychol. Bull.*, **40**, 1079–1091. <https://doi.org/10.1177/0146167214535812>.
- Hess, J. and Fila, N. (2016a) The Development and Growth of Empathy Among Engineering Students. *American Society for Engineering Education Annual Conference*, New Orleans, LA.
- Hess, J. L. and Fila, N. D. (2016b) The manifestation of empathy within design: findings from a service-learning course. *Co Design*, **12**, 93–111. <https://doi.org/10.1080/15710882.2015.1135243>.
- Hess, J. L., Strobel, J., Pan, R. C. and Wachter Morris, C. A. (2017) Insights from industry: a quantitative analysis of engineers' perceptions of empathy and care within their practice. *Eur. J. Eng. Educ.*, **42**, 1128–1153. <https://doi.org/10.1080/03043797.2016.1267717>.
- Heylighen, A. and Dong, A. (2019) To empathise or not to empathise? Empathy and its limits in design. *Des. Stud.*, **65**, 107–124. <https://doi.org/10.1016/j.destud.2019.10.007>.
- Hodges, S. D. and Biswas-Diener, R. (2007) Balancing the Empathy Expense Account: Strategies for Regulating Empathic Response. In *Empathy in Mental Illness*, pp. 389–407. Cambridge University Press, Cambridge.
- Hogan, R. (1969) Development of an empathy scale. *J. Consult. Clin. Psychol.*, **33**, 307–316. <https://doi.org/10.1037/h0027580>.
- Hojat, M. (2016) *Empathy in Health Professions, Education, and Patient Care*. Springer Science Business Media, LLC, NY.
- Hojat, M., Mangione, S., Nasca, T. J., Cohen, M. J. M., Gonnella, J. S., Erdmann, J. B., Veloski, J. and Magee, M. (2001) The Jefferson scale of physician empathy: development and preliminary psychometric data. *Educ. Psychol. Meas.*, **61**, 349–365. <https://doi.org/10.1177/00131640121971158>.
- Holt, M. (2011) The limits of empathy: utopianism, absorption and theatricality in design. *Des. J.*, **14**, 151–164. <https://doi.org/10.2752/175630611X12984592779926>.
- Howard, M. C. (2016) A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *International Journal of Human-Computer Interaction*, **32**, 51–62. <https://doi.org/10.1080/10447318.2015.1087664>.
- Jolliffe, D. and Farrington, D. P. (2006) Development and validation of the basic empathy scale. *J. Adolesc.*, **29**, 589–611. <https://doi.org/10.1016/j.adolescence.2005.08.010>.
- Jöreskog, K. G. and Sörbom, D. (1993) *LISREL 8: Structural equation modeling with the SIMPLIS command language*, pp. xvi, 202. Lawrence Erlbaum Associates, Inc., Chicago, IL.
- Kashdan, T. B., Stikma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J. and Lazarus, R. (2018) The five-dimensional curiosity scale: capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *J. Res. Pers.*, **73**, 130–149. <https://doi.org/10.1016/j.jrp.2017.11.011>.
- Kashdan, T. B., Disabato, D. J., Goodman, F. R. and McKnight, P. E. (2020) The five-dimensional curiosity scale revised (SDCR): briefer subscales while separating overt and covert social curiosity. *Personal. Individ. Differ.*, **157**, 109836. <https://doi.org/10.1016/j.paid.2020.109836>.
- Kline, P. (2000) *A Psychometrics Primer*. Free Association Books, London, UK.
- Kline, R. B. (2011) *Principles and practice of structural equation modeling* (3rd). Guilford Press, New York.
- Konrath, S., Meier, B. P. and Bushman, B. J. (2014) Development and validation of the single item narcissism scale (SINS). *PLoS One*, **9**, e103469, 1–15. <https://doi.org/10.1371/journal.pone.0103469>.
- Konrath, S., Meier, B. P. and Bushman, B. J. (2018) Development and validation of the single item trait empathy scale (SITES). *J. Res. Pers.*, **73**, 111–122. <https://doi.org/10.1016/j.jrp.2017.11.009>.
- Koskinen, I., Battarbee, K. and Mattelmäki, T. (2003) *Empathic Design - User Experience in Product Design*. IT Press, Finland.
- Kouprie, M. and Sleeswijk Visser, F. (2009) A framework for empathy in design: stepping into and out of the User's life. *J. Eng. Des.*, **20**, 437–448. <https://doi.org/10.1080/09544820902875033>.
- Krumpal, I. (2013) Determinants of social desirability Bias in sensitive surveys: a literature review. *Qual. Quant.*, **47**, 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>.

- Kulas, J. T. and Stachowski, A. A. (2013) Respondent rationale for neither agreeing nor disagreeing: person and item contributors to middle category endorsement intent on Likert personality indicators. *J. Res. Pers.*, **47**, 254–262. <https://doi.org/10.1016/j.jrp.2013.01.014>.
- Kuusisto, A. and Päällysaho, S. (2008) Customer role in service production and innovation – looking for directions for future research. In *Tutkimusraportti – Research Report 195 - Lappeenranta University of Technology - Faculty of Technology Management - Department of Industrial Management*, p. 60.
- Kyriazos, T. A. and Stalikas, A. (2018) Applied psychometrics: the steps of scale development and standardization process. *Psychology*, **09**, 2531–2560. <https://doi.org/10.4236/psych.2018.911145>.
- Lallemant, C., Lauret, J. and Drouet, L. (2022) Physical Journey Maps: Staging Users' Experiences to Increase Stakeholders' Empathy towards Users. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 344, 1–7. <https://doi.org/10.1145/3491101.3519630>.
- Lee, Y. K., Cho, W. I., Bae, S., Choi, H., Park, J., Kim, N. S. and Hahn, S. (2022a) “Feels Like I’ve Known You Forever”: Empathy and Self-Awareness in Human Open-Domain Dialogs. In Culbertson, J., Perfors, A., Rabagliati, H. and Ramenzoni, V. (eds), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 44. Retrieved from <https://escholarship.org/uc/item/0rx675tm>.
- Lee, J. J., Yap, C. and Roto, V. (2022b) How HCI Adopts Service Design: Unpacking current perceptions and scopes of service design in HCI and identifying future opportunities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 530, 1–14. <https://doi.org/10.1145/3491102.3502128>.
- Leonard, D. A. and Rayport, J. F. (1997) Spark innovation through empathic design. *Harv. Bus. Rev.* **75**, 102–113.
- Li, J., Surma-aho, A. and Hölttä-Otto, K. (2021) Measuring Designers' Empathic Understanding of Users by a Quick Empathic Accuracy (QEA). In *33rd International Conference on Design Theory and Methodology (DTM)* Article V006T06A027 (*Proceedings of the ASME Design Engineering Technical Conference*; Vol. 6). American Society of Mechanical Engineers. <https://doi.org/10.1115/DETC2021-69407>.
- Lietz, C. A., Gerdes, K. E., Sun, F., Geiger, J. M., Wagaman, M. A. and Segal, E. A. (2011) The empathy assessment index (EAI): a confirmatory factor analysis of a multidimensional model of empathy. *J. Soc. Soc. Work Res.*, **2**, 104–124. <https://doi.org/10.5243/jsswr.2011.6>.
- Liu, A., Sosik, V. S. and Singh, K. (2018) Building Empathy: Scaling User Research for Organizational Impact. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA. Paper CS03, 1–7. <https://doi.org/10.1145/3170427.3174352>.
- Maibom, H. L. (2017) *The Routledge Handbook of Philosophy of Empathy*. Routledge, NY, USA.
- Marandi, E. and Harris, J. (2010) The impact of perceived service provider empathy on customer loyalty: some observations from the health and fitness sector. *Manag. Leis.*, **15**, 214–227. <https://doi.org/10.1080/13606719.2010.483832>.
- Mattelmäki, T., Vaajakallio, K. and Koskinen, I. (2014) What happened to empathic design? *Des. Issues*, **30**, 67–77. https://doi.org/10.1162/DESI_a_00249.
- McBane, D. A. (1995) Empathy and the salesperson: a multidimensional perspective. *Psychol. Mark.*, **12**, 349–370. <https://doi.org/10.1002/mar.4220120409>.
- McDonagh, D. (2010) Rethinking design thinking: empathy supporting innovation. *Australasian Medical Journal*, **458-464**, 458–464. <https://doi.org/10.4066/AMJ.2010.391>.
- Ng, I., Maull, R. and Smith, L. (2009) Embedding the New Discipline of Service Science. In Demirkan, H., Spohrer, J. C. and Krishna, V. (eds), *The Science of Service Systems*, pp. 13–35. Springer, US. https://doi.org/10.1007/978-1-4419-8270-4_2.
- Ngo, L. V., Nguyen, T. N. Q., Tran, N. T. and Paramita, W. (2020) It takes two to tango: the role of customer empathy and resources to improve the efficacy of frontline employee empathy. *J. Retail. Consum. Serv.*, **56**, 102141–102141. <https://doi.org/10.1016/j.jretconser.2020.102141>.
- Nguyen, T. N. Q., Ngo, L. V. and Surachartkumtonkun, J. (2019) When do-good meets empathy and mindfulness. *J. Retail. Consum. Serv.*, **50**, 22–29. <https://doi.org/10.1016/j.jretconser.2019.03.020>.
- Nunnally, J. C. (1978) *Psychometric Theory*. McGraw-Hill, New York, USA.
- O'Connor, B. P. (2000) SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrum. Comput.*, **32**, 396–402. <https://doi.org/10.3758/BF03200807>.
- Parasuraman, A. P., Zeithaml, V. and Berry, L. (1988) SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *J. Retail.*, **64**, 12–40.
- Paulhus, D. L. (1991) Measurement and Control of Response Bias. In Robinson, J. P., Shaver, P. R., and Wrightsman, L. S. (eds), *Measures of personality and social psychological attitudes*, pp. 17–59. Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>.
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z. and Damer, E. (2022) Data quality of platforms and panels for online behavioral research. *Behav. Res. Methods*, **54**, 2618–2620. <https://doi.org/10.3758/s13428-022-01909-1>.
- Petermann, F. and Gust, N. (2016) *Inventar zur Erfassung emotionaler Kompetenzen bei Drei- bis Sechsjährigen (EMK 3–6)*, Testzentrale. Hogrefe, Göttingen.
- Postma, C. E., Zwartkruis-Pelgrim, E., Daemen, E. and Du, J. (2012) Challenges of doing empathic design: Experiences from Industry. *International Journal of Design*, **6**, 59–70.
- Pratiwi, H., Kristina, S. A., Widayanti, A. W. and Prabandari, Y. S. (2022) Educational interventions to improve the empathy of pharmacy students toward. *Geriatrics*, **33**, 174–185. <https://doi.org/10.22146/ijp.3837>.
- Preston, S. D. (2007) A perception-action model for empathy. In Farrow, T. and Woodruff, P. (eds), *Empathy in mental illness*, pp. 428–447. Cambridge University Press. <https://doi.org/10.1017/CBO9780511543753.024>.
- Preston, S. D. and De Waal, F. B. (2002) Empathy: its ultimate and proximate bases. *Behav. Brain Sci.*, **25**, 1–20. <https://doi.org/10.1017/S0140525X02000018>.
- Raijmakers, B., Gaver, W. W. and Bishay, J. (2006) Design documentaries: Inspiring design research through documentary film. In *Proceedings of the 6th conference on Designing Interactive systems (DIS '06)*, pp. 229–238. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/1142405.1142441>.
- van Rijn, H., Sleswijk Visser, F., Stappers, P. J. and Ozakar, A. D. (2011) Achieving empathy with users: the effects of different sources of information. *CoDesign*, **7**, 65–77. <https://doi.org/10.1080/15710882.2011.609889>.
- Rogers, C. R. (1959) A Theory of Therapy, Personality, and Interpersonal Relationships: As Developed in the Client-centered Framework. In Koch, S. (ed.), *Psychology: A Study of Science (Vol. 3)*. McGraw-Hill, New York.

- Rojas, C., Zuccarelli, E., Chin, A., Patekar, G., Esquivel, D. and Maes, P. (2022) Towards Enhancing Empathy Through Emotion Augmented Remote Communication. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA. Article 454, 1–9. <https://doi.org/10.1145/3491101.3519797>.
- Roto, V., Mitchell, V., Cockbill, S., Lee, J.-J., Law, L.-C. and Zimmerman, J. (2020) Introduction to Service Design for UX Designers. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA. Article 148, 1–3. <https://doi.org/10.1145/3411763.3444999>.
- Saco, R. M. and Goncalves, A. P. (2008) Service design: an appraisal. *Design Management Review*, **19**, 10–19. <https://doi.org/10.1111/j.1948-7169.2008.tb00101.x>.
- Salmi, A., Li, J. and Holttä-Otto, K. (2022) Facial Expression Recognition as a Measure of User-Designer Empathy. *Proceedings of the ASME 2022 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 6: 34th International Conference on Design Theory and Methodology (DTM)*. St. Louis, Missouri, USA. August 14–17, 2022. V006T06A022. ASME. <https://doi.org/10.1115/DETC2022-88924>.
- Sanders, E. B. N. and Stappers, P. J. (2008) Co-creation and the new landscapes of design. *CoDesign*, **4**, 5–18. <https://doi.org/10.1080/15710880701875068>.
- Sanders, L. and Stappers, P. J. (2012) *Convivial Toolbox: Generative Research for the Front End of Design (1st)*. BIS Publishers B.V, Amsterdam.
- Sandman, H., Meguid, T., and Levänen, J. (2020) Unboxing empathy: Reflecting on architectural design for maternal health. *CoDesign*, 1–19. <https://doi.org/10.1080/15710882.2020.1833935>.
- Shen, L. (2010) On a scale of state empathy during message processing. *West. J. Commun.*, **74** October–December 2010, 504–524. <https://doi.org/10.1080/10570314.2010.512278>.
- Simms, L. J., Zelazny, K., Williams, T. F. and Bernstein, L. (2019) Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychol. Assess.*, **31**, 557–566. <https://doi.org/10.1037/pas0000648>.
- Sleeswijk, Visser, F. and Kouprie, M. (2008) Stimulating Empathy in Ideation Workshops. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008, PDC '08*, pp. 174–177. Indiana University, USA.
- Smeenk, W. (2019) *Navigating Empathy: Empathic Formation in Co-design*. Doctoral dissertation. Technische Universiteit Eindhoven, Eindhoven.
- Smeenk, W., Tomico, O. and van Turnhout, K. (2016) A systematic analysis of mixed-perspectives in empathic design: not one perspective encompasses all. *Int. J. Des.*, **10**, 31–48. <http://ijdesign.org/ojs/index.php/IJDesign/article/view/2543/738>.
- Smeenk, W., Sturm, J. and Eggen, B. (2019) A comparison of existing frameworks leading to an empathic formation compass for Co-design. *Int. J. Des.*, **13**, 53–68. <http://www.ijdesign.org/index.php/IJDesign/article/view/3406>.
- Smeenk, W., Sturm, J., Terken, J. and Eggen, B. (2018) A systematic validation of the empathic handover approach guided by five factors that Foster empathy in design. *CoDesign*, **15**, 308–328. <https://doi.org/10.1080/15710882.2018.1484490>.
- Spector, P. E. (1992) *Summated Rating Scale Construction: an Introduction*. Sage, Newbury Park, CA. <http://dx.doi.org/10.4135/9781412986038>
- Spreng, R. N., McKinnon, M. C., Mar, R. A. and Levine, B. (2009) The Toronto empathy questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J. Pers. Assess.*, **91**, 62–71. <https://doi.org/10.1080/00223890802484381>.
- Stadler, C., Janke, W. and Schmeck, K. (2004) *IVE - Inventar zur Erfassung von Impulsivität, Risikoverhalten und Empathie bei 9- bis 14-jährigen Kindern*. Hogrefe, Göttingen.
- Stein, M., Meurer, J., Boden, A. and Wulf, V. (2017) Mobility in Later Life: Appropriation of an Integrated Transportation Platform. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, pp. 5716–5729. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3025453.3025672>.
- Stickdorn, M., Hormess, M. E., Lawrence, A. and Schneider, J. (2018) *This Is Service Design Doing: Applying Service Design Thinking in the Real World*. O'Reilly Media, Sebastopol, California.
- Surma-aho, A. and Hölttä-Otto, K. (2022) Conceptualization and operationalization of empathy in design research. *Des. Stud.*, **78**, 101075. <https://doi.org/10.1016/j.destud.2021.101075>.
- Surma-aho, A. O., Björklund, T. A. and Hölttä-Otto, K. (2018) Assessing the Development of Empathy and Innovation Attitudes in a Project-based Engineering Design Course. In *2018 ASEE annual conference & exposition proceedings (ASEE annual conference & exposition proceedings)*. American Society for Engineering Education. <https://peer.asee.org/29826>.
- Surma-aho, A., Chen, C., Hölttä-Otto, K. and Yang, M. (2019) Antecedents and Outcomes of Designer Empathy: A Retrospective Interview Study. *Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 7: 31st International Conference on Design Theory and Methodology*. Anaheim, California, USA. August 18–21, 2019. V007T06A033. ASME. <https://doi.org/10.1115/DETC2019-97483>.
- Swain, S. D., Weathers, D. and Niedrich, R. W. (2008) Assessing three sources of Misresponse to reversed Likert items. *J. Mark. Res.*, **45**, 116–131. <https://doi.org/10.1509/jmkr.45.1.116>.
- Tellez, F. A. and Gonzalez-Tobon, J. (2019) Empathic design as a framework for creating meaningful experiences. In Bohemia, E., Gemser, G., Fain, N., de Bont, C. and Assoreira A. R. (eds), *Conference proceedings of the Academy for Design Innovation Management: Research Perspectives In the era of Transformations (908-918)*, 19–21 June, Academy for Design Innovation Management, London, UK.
- Tripp, C. (2013) No empathy—No service. *Design Management Review*, **24**, 58–64. <https://doi.org/10.1111/drev.10253>.
- Vargo, S. L., Maglio, P. P. and Akaka, M. A. (2008) On value and value co-creation: a service systems and service logic perspective. *Eur. Manag. J.*, **26**, 145–152. <https://doi.org/10.1016/j.emj.2008.04.003>.
- Vasques, R., Koria, M. and Santos, M. C. L. (2022) Building Empathy in a Digital Business through Love and Break-up Letters. In Sarantu, M. and Miettinen, S. (eds), *Empathy and Business Transformation*, p. 13, Routledge, New York.
- Walther, J., Miller, S. E. and Sochacka, N. W. (2017) A model of empathy in engineering as a Core skill, practice orientation, and professional way of being. *J. Eng. Educ.*, **106**, 123–148. <https://doi.org/10.1002/jee.20159>.
- Walther, J., Brewer, M. A., Sochacka, N. W. and Miller, S. E. (2020) Empathy and engineering formation. *J. Eng. Educ.*, **109**, 11–33. <https://doi.org/10.1002/jee.20301>.
- Wambsganss, T., Niklaus, C., Söllner, M., Handschuh, S. and Leimeister, J. M. (2021) Supporting Cognitive and Emotional Empathic Writing of Students. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Online, pp. 4063–4077. August 1–6.

- Wambsganss, T., Soellner, M., Koedinger, K. R. and Leimeister, J. M. (2022) Adaptive Empathy Learning Support in Peer Review Scenarios. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 227, 1–17. <https://doi.org/10.1145/3491102.3517740>.
- Weber, F., Schween, S., Wambsganss, T. and Soellner, M. (2022) A speech-based empathy training system - initial design insights, ECIS 2022. *Proceedings Research-in-Progress Papers*, **44**. https://aisel.aisnet.org/ecis2022_rip/44.
- Weijters, B. and Baumgartner, H. (2012) Misresponse to reversed and negated items in surveys: a review. *J. Mark. Res.*, **49**, 737–747. <https://doi.org/10.1509/jmr.11.0368>.
- Wieseke, J., Geigenmüller, A. and Kraus, F. (2012) On the role of empathy in customer-employee interactions. *J. Serv. Res.*, **15**, 316–331. <https://doi.org/10.1177/1094670512439743>.
- Wright, P. and McCarthy, J. (2008) Empathy and Experience in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pp. 637–646. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/1357054.1357156>.
- Zeng, B., Wen, H. and Zhang, J. (2020) How does the valence of wording affect features of aScale? The method effects in the undergraduate learning burnout scale. *Front. Psychol.*, **11**. <https://doi.org/10.3389/fpsyg.2020.585179>.

Appendices

Appendix A. 16-item experimental version of EMPA-D (not validated)

Item code	Item
	<i>Emotional interest (EI)</i>
EI1	I am interested to learn about users' experiences and needs
EI2	I imagine how users think, feel or behave in different situations
EI3	I am curious about users' experiences and needs
EI4	I want to learn about users' experiences and opinions about the service
	<i>Sensitivity/Immersion (S)</i>
S1	I am sensitive to the experiences of users
S2	I observe without judging how users experience the service
S3	When thinking about the service, I take the users' point of reference
S4	I immerse myself in the user's world
	<i>Personal experience (PE)</i>
PE1	When thinking about the service, I consider and reflect on my own experiences and feelings
PE2	The experiences and feelings of users resonate with my own
PE3	I understand the users' experiences because I know how it feels
PE4	I compare users' experiences with the ones of people I know
	<i>Self-awareness (SA)</i>
SA1	I imagine how I would feel and think if I were a user rather than an employee
SA2	I am aware that my experiences as an employee are different from the ones of users
SA3	I realize that there are similarities and differences between my experiences and the ones of users
SA4	I understand why users perceive things differently than I do as an employee

Appendix B. 9-item EMPA-D version in Study 1

Item code	Item
	<i>Emotional interest (EI)</i>
EI1	I am interested to learn about users' experiences and needs
EI3	I am curious about users' experiences and needs
EI4	I want to learn about users' experiences and opinions about the service
	<i>Personal experience (PE)</i>
PE2	The experiences and feelings of users resonate with my own
PE3	I understand the users' experiences because I know how it feels
PE4	I compare users' experiences with the ones of people I know
	<i>Self-awareness (SA)</i>
SA2	I am aware that my experiences as an employee are different from the ones of users
SA3	I realize that there are similarities and differences between my experiences and the ones of users
SA4	I understand why users perceive things differently than I do as an employee

Appendix C. Experimental version of *Perspective-taking* dimension administered (not validated)

Initial item code	Item
PT1	I imagine how things look from the users' perspective
PT2	I am consulting data about users whenever it is accessible to me
PT3	As an employee, I try to find out what the users' needs are
PT4	When I'm frustrated by users, I try to 'put myself in their shoes'
PT5	I take action to view things from the users' perspective
PT6	I challenge my perspective with the one of the users
PT7	I try to get out of my comfort zone to 'put myself in the users' shoes'
PT8	Even when I'm convinced of something, I try to listen to users' arguments
PT9	I am actively listening to users' experiences to better understand their perspectives
PT10	I pay attention to users' complaints and testimonials about their service experience
PT11	I do my job with the users in mind
PT12	I try to think like the users
PT13	When doing my job, I try to 'put myself in the users' shoes'
PT14	When users criticize the service, I try to 'put myself in their shoes'
PT15	I imagine what it is like to be in the users' shoes

Appendix D. Final version of the 11-item Empathy in Design Scale (EMPA-D)

Instructions: The table below includes statements related to your professional context. By ‘users’ we mean people/customers who use your organization’s products or services. Please use the 7-point scale* to indicate the degree to which these statements accurately describe you or not. Respond spontaneously: there are no right or wrong answers, only your perspective matters.

Item code	Item	Initial item code
	<i>Emotional interest/Perspective-taking (EIPT)</i>	
EIPT-1	I am interested to learn about users' experiences and needs	EI1
EIPT-2	I am curious about users' experiences and needs	EI3
EIPT-3	I want to learn about users' experiences and opinions about the service	EI4
EIPT-4	As an employee, I try to find out what the users' needs are	PT3
EIPT-5	I take action to view things from the users' perspective	PT5
EIPT-6	I am actively listening to users' experiences to better understand their perspectives	PT9
	<i>Personal experience (PE)</i>	
PE-1	The experiences and feelings of users resonate with my own	PE2
PE-2	I understand the users' experiences because I know how it feels	PE3
	<i>Self-awareness (SA)</i>	
SA-1	I am aware that my experiences as an employee are different from the ones of users	SA2
SA-2	I realize that there are similarities and differences between my experiences and the ones of users	SA3
SA-3	I understand why users perceive things differently than I do as an employee	SA4

*The 7-point of the scale are: 1/Does not describe me at all 2/Barely describes me 3/Somewhat describes me 4/Moderately describes me 5/Generally describes me 6/Mostly describes me 7/Completely describes me.