This is a self-archived – parallel published version of this article in the publication archive of the University of Vaasa. It might differ from the original.

How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment

Author(s): Jansen, Bernard J.; Jung, Soon-Gyo; Nielsen, Lene; Guan, Kathleen W.; Salminen, Joni

Title: How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment

Year: 2022

Versio: Published version

Please cite the original version:

How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment

Bernard J. Jansen[1,*], Soon-Gyo Jung[2], Lene Nielsen[3], Kathleen W. Guan[4], Joni Salminen[5]

[1,*]Qatar Computing Research Institute, Qatar, jjansen@acm.org
[2]Qatar Computing Research Institute, Qatar, sjung@hbku.edu.qa
[3]IT University of Copenhagen, Denmark, lene@itu.dk
[4]University College London, United Kingdom, kathleen.guan.20@ucl.ac.uk
[5]Qatar Computing Research Institute, Qatar; Turku School of Economics at the University of Turku, Finland; School of Marketing and Communication, University of Vaasa, Finland, joni.salminen@utu.fi

Abstract

Background: *Personas are a technique for enhanced understanding of users and customers to improve the user-centered design of systems and products. Their creation can be categorized using three persona creation methodologies: Qualitative, Quantitative, and Mixed Methods. Despite the apparent differences in these methodologies, no previous review has systemically compared and contrasted the strengths and weaknesses of each of these methodologies for persona development.*

Method: *This manuscript maps and navigates persona literature to identify the benefits and challenges of these three persona creation methodologies. Furthermore, the strategies and opportunities of the different methodologies are presented.*

Results: *The results summarize the strengths and weaknesses of each of the three principal persona creation methodologies and offer suggestions of the benefits of their employment.*

Conclusion: *In conclusion, we offer insights into the construction and usage of personas for practitioners and researchers, and we propose a framework to determine which persona creation methodology is most suitable for a given context.*

Keywords: Algorithmically-Generated Personas, Persona Analytics, Persona Science.

Introduction

Digital transformation research is decidedly interdisciplinary, with offerings from domains that include HCI, Information System Sciences (ISS), and Management of Information Systems (MIS). The key to this innovation is user understanding and derived user insights to fashion the most optimal UX possible. Many methods for user understanding to design user-friendly products have been proposed within the field of Human-Computer Interaction (HCI) in particular (Baxter et al., 2015), and these methods have been employed in a range of fields relevant to design, such as digital product innovation, marketing, e-commerce, digital advertising, medical informatics, and cybersecurity (Dupree et al., 2016; LeRouge et al., 2013; Poulain & Tarissan, 2020; Salminen et al., 2019).

To create an optimal UX, choosing a particular method for user understanding depends on the user data available for an explicit task, situation, or context (Y. Li et al., 2021). The application of user segmentation in design focuses on improving user experience (UX) as a means to facilitate design processes, enhance the creation of innovations and novel business models, and advance business outcomes (Gilsing et al., 2021; Ho & Hsu, 2022; Nambisan et al., 2017; Vallet et al., 2021). In some circumstances, stakeholders require precise user data, while in other circumstances, a summary or a broad overview of users is adequate. The personas method is one such user understanding concept that supports both broad and granular user understanding.

As a well-known user segmentation technique, personas are beneficial for value creation in digital innovation (Hinings et al., 2018) and benefit information system design (Hassandoust et al., 2016). Initially, Cooper (2004) presented the notion of personas in HCI to better understand and communicate diverse user types' goals, challenges, and needs. These personas are presented as fictitious user profiles that portray representative (or significant/impactful) user groups. Personas also make visible an explicit user category into a descriptive sketch that can be accessed and comprehended even by decision makers who lack advanced user analytics skills, interests, or abilities. In short, personas of users can, among others, make the users seem real, helping designers throughout the user-centered design process. Since Cooper's seminal textbook, designers and software developers have employed personas, and other domains have incorporated personas, including, but not limited to, cybersecurity, marketing, video game studies, and health informatics (Mulder & Yaar, 2006; Nielsen, 2013; Pruitt & Adlin, 2010).

The commonly argued benefit of employing personas comes from succinctly presenting user information in a natural form that end users of personas can communicate with minute effort (Holtzblatt et al., 2005) both within an organization and with external stakeholders (Matthews et al., 2012). Personas merge design activities for the core user segments without direct interaction with the targeted user (Floyd et al., 2008). Personas are often based on preexisting user information in various forms, from small samples of qualitative interviews to big data analytics in various data shapes. Personas also meld user research data into easily useable and understandable representations of users. In principle, personas deliver an appealing description of the users' wants and needs (H. Li et al., 2022) by presenting a description of a human being that is more empathetic than using mere numbers about users (Goodwin & Cooper, 2009; Hill et al., 2017). Ideally, the cast (i.e., set) of personas become shared and aligned mental models of the users that stakeholders use and consult for design decisions (Nielsen, 2013) concerning specific user types (Cooper, 2004). Personas thus offer a human perspective for decision-makers to converse backgrounds and experiences that may be different from their own (Miaskiewicz & Kozar, 2011).

In the literature, three core persona creation approaches exist: *Qualitative, Quantitative, and Mixed Methods* (Jansen et al., 2021; Tu et al., 2010). Each of these creative approaches comprises different types of data collection, processing, and analysis (including qualitative,

quantitative, and algorithmic), such as meaning categorization via empathy diagrams, hierarchical clustering, k-mean clustering, latent semantic analysis, weighted graphs, regression analysis, and so on (Hazrati & Ricci, 2022; Zhu et al., 2019). However, the boundaries among these three creation methodologies are blurred, so there can be overlap in their use in practice.

This variety of persona creation methodologies may be perplexing and problematic, especially for novice persona developers. Persona approaches vary tremendously, from unstructured qualitative methods to complicated data science algorithms (Salminen, Guan, et al., 2020). A rudimentary foundation of knowledge is necessary for choosing the suitable persona creation methodology for a given setting, scenario, or tasking.

This manuscript presents a synthesis of the strengths and weaknesses of the three general persona creation methodologies that assist in this purpose by building on prior work in the field (Jansen et al., 2021; Tu et al., 2010). We aim to inform and inspire both novices (by providing foundational information about persona creation methodologies) and experienced persona creators (by helping them to discover approaches outside their comfort zone).

The research questions (RQs), therefore, are:

- **RQ1:** What are the (a) strengths and (b) weaknesses of each of the three persona creation methodologies?
- **RQ2:** What is the optimal context for employing each methodology?
- **RQ3:** What are the significant opportunities of personas for innovative user-centered design within organizations?

Different persona creation methodologies have been deliberated in a range of previous HCI studies (Nielsen, 2004, 2013; Pruitt & Adlin, 2010; Pruitt & Grudin, 2003). Nevertheless, to our knowledge, no prior systematic review has focused on analyzing the strengths and weaknesses of the three approaches and their implementation, and such a comparative analysis is one of our primary goals. In addition, we summarize the strategic opportunities for using personas to enhance user understanding. Personas are also used differently in varying domains, e.g., design personas might traditionally highlight empathetic user qualities while marketers and new product developers might emphasize quantitative information for purposes such as targeting or market share capturing. These considerations make the assessment of strengths and weaknesses somewhat dependent on the immediate context in which the persona is deployed. Here, we aim to adopt a general view of personas, pointing out relevant aspects across domains.

Three Principal Methods of Persona Creation

Qualitative persona creation (QUAL) consists of data collection and analysis that are usually manual. Examples of qualitative data collection methods are focus groups, user interviews, and (sometimes) surveys (Miaskiewicz et al., 2008; Vosbergen et al., 2015), although surveys can also be a method for quantitative data collection. The user data is often descriptive and unstructured, e.g., texts and interviews (Dupree et al., 2016; Ford et al., 2017; Huh et al., 2016). Demographic data can be helpful for use cases such as look-alike analysis (Duvvuri, 2021). Qualitative analysis approaches typically employed are interpretative, such as open and axial coding (H. Guo & Razikin, 2015; Mesgari et al., 2015). Although one can create assumption-based personas without any data (Matthews et al., 2012; Seiden & Gothelf, 2003), personas are more commonly produced based on at least some user data or data pooled with fictitious attributes. When actual user data is employed, personas are classically generated using ethnographic means and/or interviews with users (Cooper, 2004; Goodwin & Cooper,

2009; Pruitt & Grudin, 2003), which typically means that the personas rely on sample sizes that may not enable applying statistical analysis.

Quantitative persona creation (QUANT), however, typically encompasses methods of automatic data collection and application of statistical analysis and data science algorithms or the use of quantitative data collection methods (e.g., online surveys). An example of such an automated approach would be using Application Program Interfaces (APIs) (Cleland-Huang et al., 2013; Wang et al., 2018) for data collection. The data for QUANT is customarily structured, e.g., by frequency of user engagement or responses and demographic grouping. Applications of QUANT methods include factorization, regression, and clustering, among others (An, Kwak, Jung, et al., 2018; An, Kwak, Salminen, et al., 2018; Walsh et al., 2021). While creating data-driven personas using large-scale behavioral data has been limited (McGinn & Kotamraju, 2008), the use of QUANT methods has increased over the years. This movement is driven by the growing availability of online user data (Deng et al., 2020; Xie et al., 2020) that is easily accessible and automatic user segmentation algorithms (Salminen, Guan, et al., 2020) that are reasonably simple to implement on such large volumes of data. QUANT can also employ data collection methods, such as surveys executed on large sample sizes.

The separation of persona creation methodologies into distinct QUAL and QUANT is commonplace in many fields. However, there is also the desire to combine the two, resulting in a **mixed-methods approach (MIXED)** (McGinn & Kotamraju, 2008; Tan et al., 2021; Tu et al., 2010). The premise of the MIXED methodology is that both QUAL and QUANT methodologies are well-suited to be used in combination (Sun et al., 2021) and can be used mutually to produce more comprehensive (well-rounded) personas than neither of the sole approaches could achieve on its own (McGinn & Kotamraju, 2008).

Personas derived from the MIXED methodology are frequently recommended in the research literature (Pruitt & Grudin, 2003), while QUANT methods generally are presented alongside QUAL data and analysis in writing and evaluating the persona profiles (Salminen, Guan, et al., 2020). Finally, a QUAL persona might also contain quantitatively derived data, although typically in limited quantities. So, we again acknowledge that the distinction among these three persona creation approaches is not absolute, and some overlap exists.

The steps for creating personas using the three principal persona creation methodologies are outlined in Table 1.

Table 1 – Persona creation steps. The first and last steps are identical for all three methodologies.			
	Qualitative	**Quantitative**	**Mixed-Methods**
Step 1: Choose the purpose	Determine the purpose(s) for which the personas will be used.		
Step 2: Collect the data	Manual data gathering methods such as interviews, focus groups, and surveys (usually with small sample sizes) are used to obtain user data.	The user data is often gathered automatically from analytics platforms or other sources such as online surveys and Customer Relationship Management (CRM) systems. Data will usually comprise both demographics and digital behavioral attributes.	The user data is typically gathered via automatic means from analytics platforms or other sources such as surveys and CRM systems. After identifying data shortfalls, enrich the data via qualitative methods (or vice versa, with qualitative data being collected first followed by quantitative).
Step 3: Analyze the data	Analyze user data for demographic or behavioral characteristics, generally using conventional qualitative approaches (e.g., grounded theory).	Analyze user data for demographic or behavioral patterns, generally using quantitative or computational approaches.	Using a combination of qualitative, quantitative, and/or computational methodologies, analyze the user data to find demographic or behavioral patterns.
Step 4: Identify standard users	From the results of data analysis, identify the key user segments.	The quantitative or algorithmic method will yield an exact number of relevant user segments.	From the results of data analysis, identify the key user segments. The core segments commonly result from the algorithmic method, but other segments may emerge with the data enrichment.
Step 5: Create the cast of persona profiles	Enhance user segments with name, picture, topics of interest, quotes, etc., to create persona profiles for each. The result is a complete cast of personas.		

Collection of Prior Literature and Analysis of Prior Work

To conduct our prior review of the three persona creation approaches, we first collected 74 research articles containing both theoretical and experimental persona articles. The articles were selected based on their relevance to persona creation, and they were all peer-reviewed. We consulted relevant databases (ACM Digital Library, Google Scholar) to search for the articles, including venues that often publish persona research, such as UX-focused conferences (e.g., CHI, INTERACT). We then analyzed the material by (a) reading each of the 74 research articles, (b) comparing the persona-related aspects of each, and (c)

synthesizing the research findings systematically, leaving a complete systematic review for other research.

We then performed an analysis by having each researcher read the essential sections of the publications independently. The "codes" (Glaser & Strauss, 2009) pertaining to each persona creation approach, as well as its strengths and weaknesses, were recorded and analyzed in a codebook (Strengths-Qualitative, Weaknesses-Qualitative, etc.). A code is an inductive unit of observation in the qualitative research paradigm that describes a relevant trend in the data, applicable to the research purpose and data in terms of the articles. 'COMPLEXITY' is an example of a code (see Section 4.1 for further discussion). The researchers worked together to create the codes, with each researcher providing a code to each component of the articles. The researchers then evaluated each other's codes, resulting in the final codes list.

The subsequent sections discuss the results of this literature analysis, addressing **RQ1**. We offer <u>supporting references (SR)</u> for each code (see the boxes presenting the conclusions of the analysis).

Strengths and Weaknesses of Each Approach

Qualitative Persona (QUAL) Creation

The point of origin for personas is for software design and software development, and the advocacy for personas was initially two-fold: firstly, (a) to conjure empathy for the target user segments one designs for (Grudin, 2006), and secondly, (b) design teams incorporated the persona technique to understand user interests, needs, desires, pain-points, satisfactions, work processes, etc. For the first purpose, the qualitative persona method has, from the start, been a method that intended to answer the inquiries of why people behave and reason as they do. Qualitative methods were naturally beneficial, and inquiry methods such as ethnography provided instruments, such as contextual inquiry, interviews, and observations that support the formation of an in-depth understanding of users. Accessing in-depth insights for design areas is challenging using quantitative methods, while qualitative methods struggle with analyzing data at scale.

More specifically, the QUAL persona methodology has several strengths explaining why many persona designers prefer and like this approach. The strengths (itemized in alphabetical order) are:

- **COMPLEXITY:** Examining nuanced and multi-layered user behaviors (i.e., multifaceted user attributes). *SR:* (Blomquist & Arvola, 2002; Mulder & Yaar, 2006).

- **DEPTH:** Concentrating on a limited number of persona use cases in substantial depth. *SR:* (Adams et al., 2008; J. Guo & Yan, 2011).

- **DESCRIPTIVE:** Inductively generating a descriptive model of a user type that uses human judgment and interpretation. *SR:* (Neate et al., 2019; Nielsen, 2004).

- **EMOTIONS:** Transmitting the users' understandings, emotions, views, and beliefs as social constructs into the persona profile. *SR:* (Mulder & Yaar, 2006; Pruitt & Grudin, 2003).

- **EMPATHY:** Accessing the causal context of behaviors, needs, goals, feelings, and pain points to produce explanations as to why the persona thinks/does/acts in a certain way. *SR:* (Hisham, 2009; Nielsen, 2004; Wright & McCarthy, 2008).

- **EVALUATION:** Examining dissimilar user types and circumstances based on manual contrasting of central user attributes. *SR:* (Aljohani & Blustein, 2015; J. Guo & Yan, 2011).

- **EXPERIENCES:** Offering an understanding of users' experiences from a personal level, leveraging the creators' innate personal perceptions. *SR:* (Adams et al., 2008; Idoughi et al., 2012).

- **PERSONALIZE:** Providing manually curated anecdotes and insights to be employed within the persona profiles. *SR:* (Anvari et al., 2017; Nielsen, 2004).

- **SPECIFICITY:** Producing rich depictions of explicit user circumstances across scenarios that matter for design. *SR:* (J. Guo & Yan, 2011; Holtzblatt et al., 2005).

In contrast, the weaknesses of QUAL are:

- **BIAS:** QUAL persona profiles can be beset with human biases, stereotypes, and idiosyncrasies. *SR:* (An, Kwak, Jung, et al., 2018; Chapman & Milham, 2006).

- **EFFORT:** Manually creating persona profiles can be time-consuming and costly. *SR:* (An, Kwak, Jung, et al., 2018; Drego et al., 2010).

- **INVALID:** Qualitative personas risk having low credibility levels among persona users who prefer "hard evidence". *SR:* (An, Kwak, Jung, et al., 2018; Mesgari et al., 2015).

- **NARROW:** There may be a lack of generalizability of the created personas to other users or contexts. *SR:* (Brickey et al., 2010; Chapman et al., 2015).

- **REPRESENTATIVENESS:** Small sample sizes may result in small user segments possibly being overrepresented and the personas not representative of the entire user population. *SR:* (An, Kwak, Jung, et al., 2018; Chapman & Milham, 2006).

Example of a Persona Created via the QUAL Approach

Figure 1 presents an example of a typical persona profile generated from a QUAL methodology (Xtensio, 2021) using a standard persona profile template. As the focus is the profile elements, we defer the discussion of the crafting of the persona. In general, the example is the industry standard for a persona profile, especially one created using a QUAL methodology.
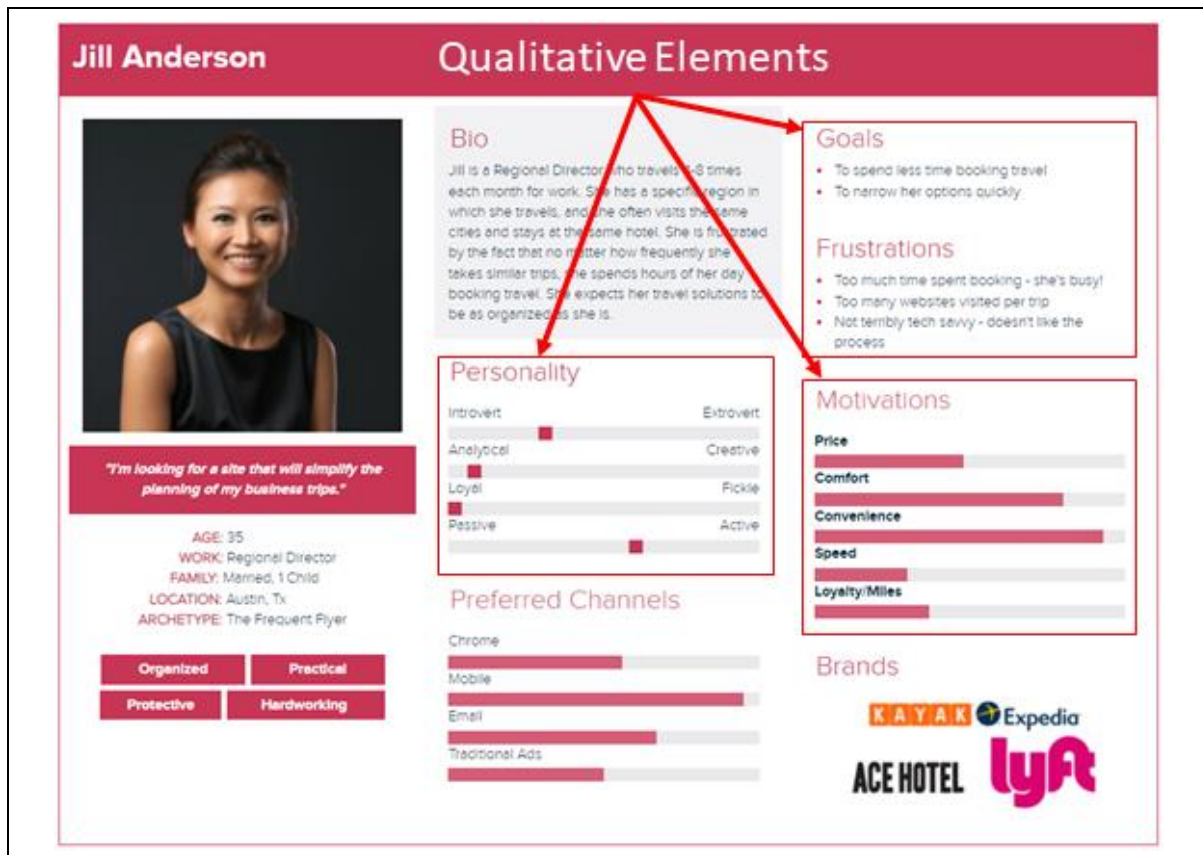
Figure 1 – Persona profile created from a QUAL approach (profile from (Xtensio, 2021) Examples of qualitative elements annotated. The bar charts are typically produced manually in a way that gives the impression of precision, but the process itself is often not precise.

As shown in Figure 1, the QUAL persona profile has the underlying user data personified into a representative person (in this case, Jill), with accompanying demographic and other related attributes. This example in Figure 1 exemplifies many of the strengths of the QUAL methodology. Not only does the profile have demographic data, but it contains a descriptive quote capturing the overall user need ("I'm looking for a site that will simplify the planning of my business trips."), along with personality attributes and preferred online channels of use. The persona profile also highlights the persona's goals, motivations, and frustrations, and the persona's preferred brands are also presented (COMPLEXITY, DEPTH, EXPERIENCES, PERSONALIZE, SPECIFICITY), with many styles of presenting such attributes. As noted in this persona profile, they are presented as sliding scales and bar charts, although the underlying data may not be as numerically precise as the graphs, scales, and charts imply.

This QUAL profile also highlights the major weakness of such a persona. Research (Chapman et al., 2008) has shown that as more and more of these in-depth attributes are added to the profile, the personas represent fewer actual individuals (REPRESENTATIVENESS, NARROW). Also, given that much of the data to construct QUAL personas is a filter through a persona designer, there are questions concerning the validity of the persona data itself (INVALID, BIAS).

Quantitative Persona (QUANT) Creation

QUANT has been used for some time with quantitative data collection methods, such as surveys and system interaction logs. However, the availability of a large volume of online user

data, both from online surveys at scale, internal sources (e.g., CRM systems), and digital analytics services (Jansen et al., 2020), combined with increasingly sophisticated algorithmic procedures and interactive persona systems (An, Kwak, Jung, et al., 2018; An, Kwak, Salminen, et al., 2018), encourage the creation of personas via quantitative data and using statistical methods.

Specifically, collecting user data via APIs has radically increased the viability of quantitative persona creation (Vecchio et al., 2018). Prominent social media services (e.g., Facebook, YouTube) and sizeable online analytics platforms (e.g., Adobe Analytics, Google Analytics) can be used to collect this online data (Dang et al., 2021; Ng & Wang, 2019).

Relying on these opportunities of data availability, the QUANT persona methodology has numerous strengths that illuminate why it has increased support and use in many persona development settings. The strengths include:

- **EVALUATION:** Permits the testing hypotheses about users developed before creating the personas. *SR:* (Brooks & Greer, 2014; J. Li et al., 2016).

- **PRECISION:** Collected data is more concise, precise, and statistically reliable than qualitative observations. *SR:* (Chapman et al., 2015; Dupree et al., 2016).

- **PRESENTATION:** User findings can be dynamically presented and updated as the data is based on large sample sizes, even up to millions of user interactions. *SR:* (An, Kwak, Jung, et al., 2018; An, Kwak, Salminen, et al., 2018).

- **REPEATABILITY:** The persona creation can be replicated using the same datasets, arriving at the same conclusions, which adds capacity to simplify user insights about many diverse segments and populations. *SR:* (An, Kwak, Jung, et al., 2018; Salminen et al., 2019).

- **SIMPLICITY:** The capacity to construct a series of data processing and analysis steps eliminates the sporadic nature of manual data analysis and enables more predictable cause-and-effect relationships. *SR:* (Kim et al., 2019; Tempelman-Kluit & Pearce, 2014).

- **SPEED:** Quicker data collection and analysis relative to the QUAL approach. *SR:* (Chu et al., 2018; Ishii et al., 2018; Mijač et al., 2018).

- **TESTING:** Persona profiles can be leveraged for quantitative predictions, such as the persona's topical interests and preferred online content. *SR:* (Miaskiewicz et al., 2008; Rahimi & Cleland-Huang, 2014).

- **VALIDITY:** Permits the validation and testing of created theories about users using observational data or experimental designs. *SR:* (Minichiello et al., 2018; Wöckl et al., 2012).

- **VOLUME:** Applicability for the studying of large numbers of users. *SR:* (An, Kwak, Salminen, et al., 2018; Watanabe et al., 2017; Zhu et al., 2019).

The QUANT approach has shortcomings, however. These include:

- **COMPLEXITY:** Gathered user data may necessitate complex algorithms for analysis that require competencies not always available within a design team. *SR:* (Dhakad et al., 2017; Holmgard et al., 2014).

- **DISCONNECTION:** Segmenting may not represent the needs and preferences of the end users, and the designer might feel disconnected from the personas due to a lack of involvement in their creation. *SR:* (Goodman-Deane et al., 2018; Hirskyj-Douglas et al., 2017).

- **OUTLIERS:** Statistical influence of the larger user segments in the persona creation process may mask the impact of outliers and marginalized user groups, thereby harming inclusive design goals. *SR:* (Tychsen & Canossa, 2008; Zhang et al., 2016).
- **TARGETED:** Created personas may represent existing users and not the desired/potential users, while also being somewhat limited in what insights can be obtained using statistical analysis. *SR:* (Bamman et al., 2013; Brickey et al., 2012; Holden et al., 2017).

Example of a Persona Created via the QUANT Approach

As a case study, we present and examine QUANT personas generated from an algorithmic persona analytics system called Automatic Persona Generation (APG) (https://persona.qcri.org), available online for inspection and is used by several clients for the generation of personas. APG uses a matrix factorization algorithm to generate QUANT personas based on large datasets containing thousands of products, hundreds of thousands of users, and tens of millions of user interactions. As such, APG represents state-of-the-art in QUANT persona generation. As APG has been discussed in prior work (Jansen et al., 2020; Salminen, Jung, et al., 2020), we only present this brief introduction; instead, we focus on the generated personas.

Using APIs to online user analytics data, APG can generate a cast of personas, with ten personas set as the default, as shown in Figure 2.
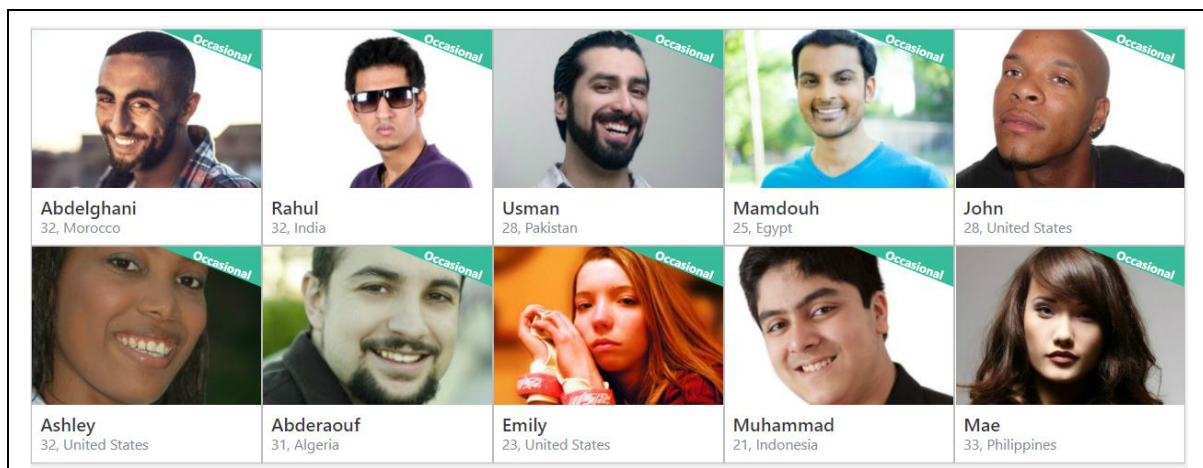

Figure 2 – Cast of QUANT personas generated from APG, an interactive persona analytics system using algorithmic methods to create personas.

Each persona in the cast is represented by meta-data of an image, name, age, and country (see Figure 2), along with their loyalty status based on engagement frequency (*Occasional* for all of the personas for the example in Figure 2). By clicking on one of the personas in the cast (for example, *Mae*), the complete persona profile is displayed, as shown in Figure 3.
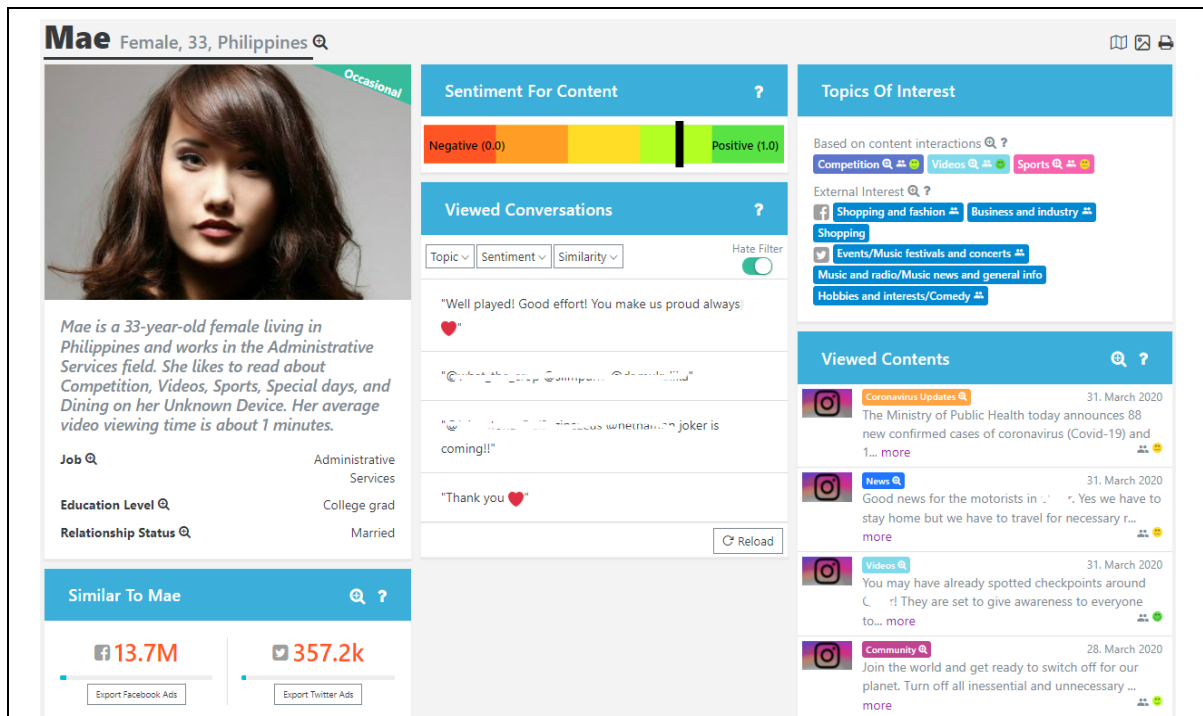
Figure 3 – Example of QUANT persona profile generated from the APG persona analytics system. Note: Identifying text in the profile is masked for the confidentiality of users and the organization.

As shown in Figure 3, the QUANT persona profile has most of the standard elements of the de facto industry standard profile (e.g., name, image, demographic information), and the QUANT profile also has many additional elements not usually found in many traditional persona profiles (e.g., estimation of segment size, sentiment, online conversations, topics of interest, and viewed content). These additional elements would be challenging to carry out using qualitative data collection methods.

This QUANT persona example illustrates many of the strengths of the QUANT approach. Once the system is built, the persona analytics system can generate personas from millions of user data points (PRESENTATION, VOLUME) within hours (SPEED). The data employed by the system is actual analytics data from real users (PRECISION). The underlying factorization algorithm identifies unique sets of user behaviors and ties these behavior sets to represent demographic groups (EVALUATION). The process is repeatable, with the system storing each generation of personas (in the case with APG, monthly) (REPEATABILITY), facilitating the comparison of personas over time. The profile has predictive components, such as sentiment (VALIDITY, TESTING).

Comparably, this QUANT example also illustrates some of the weaknesses of the QUANT approach. The cast of personas is representative of existing users and may not represent the desired users. The profiles do not contain all the information that QUAL personas have (TARGET). Also, the personas are generated algorithmically, so the organizational involvement of creating personas is not present (DISCONNECTION). Likewise, given the volume and types of data employed, the system requires various machine learning algorithms that may be beyond the skill sets of some organizations (COMPLEXITY). Moreover, the users' needs, wants, goals, and aspirations are not incorporated into the QUANT profile (TARGET) easily.

Mixed-Method Persona (MIXED) Creation

When the QUAL approach is integrated with the QUANT approach, the resulting cast of personas can often capture more varied and representative representations of user behaviors (the added value of quantitative methodology) and better construe the complexities of any condition or user group (qualitative added value) for a user population. Therefore, the MIXED approach has numerous strengths that are important to note:

- **COMPLETENESS**: Adding insights (either numeric or qualitative) that could be missed depending on a single approach. *SR:* (Dang-Pham et al., 2015; Tempelman-Kluit & Pearce, 2014).

- **DIVERSITY**: To clarify underlying meaning and user context, qualitative data (narratives) can be combined with quantitative data (numbers), resulting in socially meaningful user representations. *SR:* (Dupree et al., 2016; Hirskyj-Douglas et al., 2017).

- **FLEXIBILITY**: Produce more comprehensive knowledge essential for informing actionable insights in a specific use case. *SR:* (Hirskyj-Douglas et al., 2017; Thoma & Williams, 2009).

- **RANGE**: Allows for a broader and more thorough portrayal of the variety of information due to a variety of data collection or analysis means; not relying on methodological or dataset restrictions pertaining to a single methodological fraction. *SR:* (An, Kwak, Jung, et al., 2018; Minichiello et al., 2018).

- **RESILIENT**: Integrates the strengths of both quantitative and qualitative methods in a productive manner. *SR:* (McGinn & Kotamraju, 2008; Tu et al., 2010).

- **SUPPORT**: Better positioned profiles provide evidence of better representing users via merging findings from different individuals/mechanisms/processes, such as combining human- and algorithm-generated user analysis. *SR:* (Salminen et al., 2018; Zaugg & Ziegenfuss, 2018).

- **TESTABLE**: Profile information can be used to both produce and assess hypotheses in a progressive manner that utilizes qualitative sensemaking and quantitative verification. *SR:* (Dupree et al., 2016; Miaskiewicz & Luxmoore, 2017).

The MIXED approach can be nuanced to implement in practice, including the noted difference depending on if one starts with qualitative data and integrates quantitative data or vice versa (Miaskiewicz & Luxmoore, 2017; Mulder & Yaar, 2006). Therefore, the MIXED approach also does possess limitations, including:

- **EFFORT**: Costly and time-consuming compared with the other two methods of persona creation, as possible duplicates and the need to de-duplicate content during data collection and analysis are increased. *SR:* (Holden et al., 2017; Mijač et al., 2018).

- **IMPLEMENTATION**: Possibly challenging to implement by a single persona development team, especially when the team lacks technical or qualitative analysis skills, time, budget, or data collection resources. *SR:* (Mesgari et al., 2015; Miaskiewicz & Luxmoore, 2017).

- **INTEGRATION**: Possible problems when interpreting conflicting results into a coherent profile due to the mixture of diverse data collection sources, integration techniques, and analysis methods; the findings may be isolated and not "sit well" within the whole. *SR:* (Ford et al., 2017; Salminen et al., 2018).

- **PREPARATION**: Requires a thorough understanding of various methodologies and an understanding of how to integrate them to create cohesive personas, for which there are no definitive guidelines. *SR:* (H. Guo & Razikin, 2015; Salminen et al., 2018).

Example of a Persona Created from the MIXED Approach

Here, we present a novel example of a MIXED methodology use case. Survey2Persona uses survey data (traditionally used for QUAL personas) combined with automatic statistical analysis (employed by QUANT personas) within an operating online persona survey analytics system. As our focus is the personas, we defer detailed discussion of the system to other work.

Using data collected from a user survey (in this example, computer-generated survey responses for demonstration of the breath of the system), Survey2Persona (https://s2p.qcri.org) generates a cast of personas based on criteria selected by the persona end user (in this example, age grouping 25-34 and biological sex as female), as shown in Figure 4.
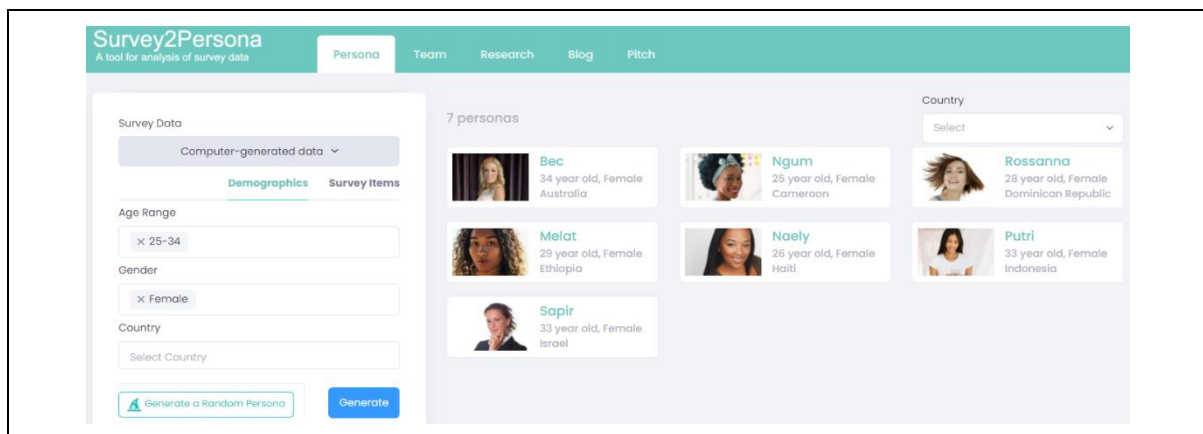

Figure 4 – Cast of MIXED personas generated from Survey2Persona, an online persona system that leverages survey data and survey data analysis.

As seen in Figure 4, the system automatically generates a cast of personas, each with associated meta-data (e.g., name, age, gender, and nationality). Clicking on one of the personas in the cast displays the completed persona profile, as shown in Figure 5 (in this example, *Bec*).
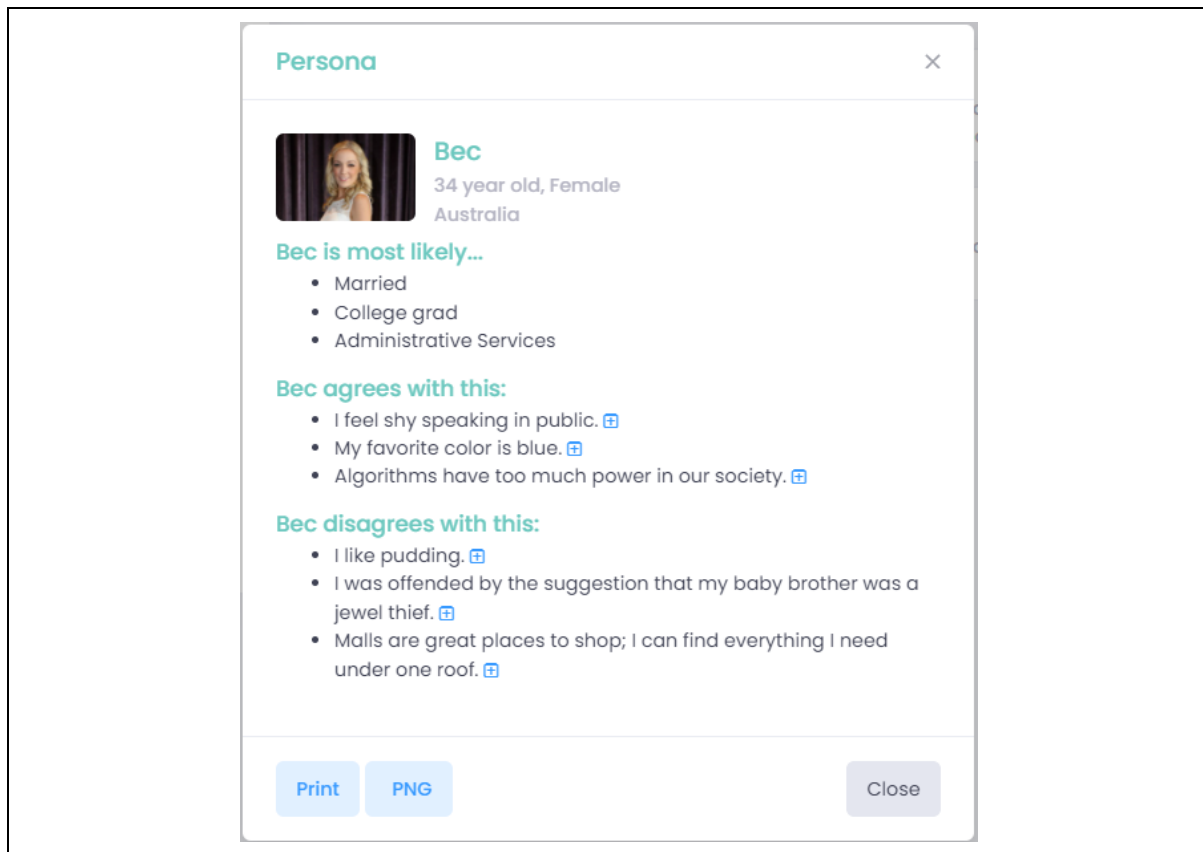
Figure 5 – Example of MIXED persona profile generated from the Survey2Persona system.

As Figure 5 shows, the MIXED persona profile has several of the standard attributes common in most persona profiles (e.g., name, image, demographic information), along with the probabilities of relationship status, occupation, and education, which are common (but calculated differently) in both QUAL and QUANT profiles. For the profile in Figure 5, these attributes are calculated in the QUANT approach using probabilities within the overall user population. The MIXED profile also has many of the textual insights, values, and opinions (Dang et al., 2021) that one sees in many QUAL profiles but are lacking in most QUANT persona profiles.

This MIXED profile illustrates many of the strengths of the MIXED approach. The profile has the standard attributes of must profiles, with a mixture of both numerical and textual information (COMPLETENESS, DIVERSITY). As such, the MIXED approach can leverage the method most appropriate to the task at hand (FLEXIBILITY, RANGE, RESILIENT). By employing data typically used for qualitative personas, the MIXED approach offers the depth of insights into the persona. By employing methods of analysis often used for the quantitative generation, the MIXED persona better represents the user data (SUPPORT) and affords various areas for testing (TESTABLE).

The MIXED example also illustrates some of the weaknesses of the MIXED approach. Integrating qualitative and quantitative data into a coherent profile takes a range of skill sets (IMPLEMENTATION) for collection and analysis. As the data comes from different sources, incorporation can be challenging (INTEGRATION), requiring considerable resources (EFFORT).

14

Discussion of Findings and Implications

Personas have a wide range of UX, system design, and digital innovation applications, making them useful in interactive marketing, design and testing, product design, user testing, and lead generation. The persona approach has intrinsic advantages over other user-centered methods in that it usually gives "cold" numbers an accessible human face and presence (Jansen et al., 2020). However, the number of persona creation approaches reported in the literature may be difficult to navigate, particularly those new to personas.

To aid in navigating the methodological plurality of persona creation, we examined the three primary methodologies of persona creation. The main implications from our findings to persona users are three-fold.

1. First, the MIXED methodology is theoretically the most robust choice, as it combines the strengths of both QUAL and QUANT methodologies, thereby addressing the weaknesses of these two methodologies.
2. Second, related to the first implications immediately above, MIXED is also the most challenging methodology to implement. It requires the most diverse set of skills from the persona creators and multiple modalities of data that may not be available without dedicated large-scale data collection efforts. Such efforts may "eat away" the efficiency gains of the QUANT methodology, thereby not effectively being the most optimal choice for situations where personas are required in a limited timeframe and/or reasonable cost.
3. Third, there is a definite demand for systematic and innovative research and development work to address the challenges we systematically mapped in this work. The scarcity of such work has resulted in much of the prior work within personas to "run in circles", so that methodological fractions (QUAL vs. QUANT) are born and remain somewhat stubbornly in their own corners, often perceiving the other options as an "alternative" (or even adversary) as opposed to being a resource or complementary asset to one's own methodology.

Three requisites seem appropriate to address the broader goal of advancing persona science and the scholarly evolution of persona creation methodologies.

First, persona scholars and practitioners must (1) *be aware of the strengths and weaknesses of each method*. For this, the Strengths, Weaknesses, Opportunities, Threats (SWOT) analysis is valuable for internalizing, as it principally speaks to **RQ1** (see Table 2). It is imperative, for example, to note that both QUAL and QUANT methodologies involve the risk of bias – while "bias" is traditionally seen as a shortcoming of human fallibility that "data" and "algorithms" can address, algorithmic bias is being increasingly recognized as being part of data-driven and algorithmic user segmentation, including quantitative persona creation. This, again, supports the notion that "no methodology is perfect" or the only one that should be advocated (which is sometimes tacitly implied by persona papers adhering to a certain methodological school of thought). Also, the "right" method depends on the context, for example, a new product design (NPD) project or a project focused on improving a system interface. In the NPD case, data from actual customers may not exist, so a QUAL methodology might be best to first identify target customers, and then a QUANT methodology using secondary data to refine these target customers. Ultimately, while adding to the complexity of understanding the persona research tradition, methodological plurality is a source of inspiration and contains the steppingstones for improving the field as a whole.

Table 2 – A SWOT analysis of the persona creation methods (QUAL, QUANT, and MIXED).			
	Qualitative	**Quantitative**	**Mixed-Methods**
Strengths	Nuanced, detailed insights	Comprehensive and testable attributes of behaviors	More refined clarifications of observed attributes
Weakness	Partial data and limited testable hypotheses	Confine insights of goals, desires, pain points	Challenging to integrate disparate data
Opportunities	Greater methods of data collection	Accessibility of online analytics data	Algorithmic methods for the analysis of qualitative data
Threats	Speedily shifting user population necessitates additional data collection rounds	Continuous alteration to APIs, platforms, and services	Diverging user segments causing contradictory insights

Secondly, persona creators ought to (2) *study the context of the persona creation methodology and end use case.* This includes considering the decision process and the context of that process of stakeholders and leveraging the personas in their design taskings. A portion of this "contextual awareness" is an understanding of how each method is companionable with the user-centric analysis context at hand. For example, the user population is small and specific, and it is not probable to collect a large user dataset for quantitative analysis – this can be the case in specific areas within health IT. Other times, the decision-makers need specific quantitative information to make decisions based on the personas.

Third, persona users must (3) *understand there is no "one best approach" to personas creation.* When resources and data allow, the MIXED approach comes the closest to this position. The MIXED approach, as recommended in existing HCI literature (Mesgari et al., 2015; Pruitt & Grudin, 2003; Salminen et al., 2018), aids in addressing questions that cannot be answered only using the QUAL or QUANT approaches. End users of personas must be aware that while incorporating data, information, and results from the QUAL and QUANT approaches into a MIXED method for persona creation, there are procedures to aid the integration. One such technique is the triangulation design model (Tashakkori & Teddie, 1998) that combines data and information from both the qualitative and quantitative collection.

Table 3 offers strategies for persona users to select the proper approach for their persona creation process. This thus addresses **RQ2**. From a comparison and analysis, these guidelines reflect seven criteria: (1) *Data*, (2) *Context*, (3) *Information*, (4) *Updatability*, (5) *Interactivity*, (6) *Timeliness*, and (7) *Economics*. The take-away is that, in the "perfect world", without limits, persona creators should generally opt for the MIXED persona creation methodology. However, myriad of practicalities – such as data availability, a lack of technical expertise, and resource limitations like GDPR, time, and money – continue to hinder this vision of perfection, and creators of persona in real life settings are frequently forced to consider trade-offs based on the prevailing conditions.

Table 3 – Guidelines to choose an appropriate persona creation approach for persona users for their task. QUANT is appropriate for 'What Questions', QUAL is appropriate for 'Why Questions', and MIXED is appropriate for both 'What' and 'Why Questions'.			
Criteria	**Select QUAL if…**	**Select QUANT if…**	**Select MIXED if…**
Data	You have access to users that are prepared to share their experiences and expectations.	You have preexisting or can gather quantitative data that designates user demographics and behaviors.	You have access to both quantitative and qualitative user information.
Context	Design and decision-making tasks necessitate an in-depth understanding of the users.	Your data can be used for design and decision-making tasks in the organization for which the personas are created. You need to know how many of each segment.	The personas need to adjust to many use cases and scenarios that are difficult to anticipate. Your tasks demand both in-depth understanding and numeric values.
Information	The information needs are on qualitative understandings (e.g., user pain points, motivations, goals)	The information needs can be content with quantitative data of user segments (e.g., system features employed).	The information needs vital in-depth personas with qualitative details (e.g., customer needs) and numerical accuracy (e.g., segment size represented by the persona).
Updatability	Personas are used for a one-time project or do not necessitate regular updating.	Personas need to be updated regularly (i.e., the attributes and demographics in the user population are rapidly changing).	If there are proper resources (time, money, expertise) to update the personas as prescribed by the changes in the user behavior.
Interactivity	There is no necessity for decision-makers to interact with the personas beyond the media of paper, presentation slides, and posters.	Decision-makers need advanced interactivity and discover personas using computer-assisted media via continually changing data.	There is an option to leverage various data types in an interactive system that provides long-term value for decision-makers.
Timeliness	If persona creation is not time-sensitive but can afford data collection using interviews and/or ethnography.	You need to generate the personas quickly without time to conduct manual data collection and analysis.	If there is no strict time limit for producing the results of quantitative and qualitative inquiry.
Economics	Adequate budget is accessible for professional interviews/ethnography.	You have preexisting resources (skills, software) that can be arranged for quantitative analysis.	If there is the budget for and the use of experts from qualitative and quantitative domains.

As there may be some overlap among the three approaches, there can also be nuanced differences in the criteria. For example, for interactivity, we refer to advanced interaction techniques that are not possible unless using digital media, such as chatting with a persona or filtering the persona's quotes by sentiment. While manual personas offer some ways of interaction (e.g., one can manually sort paper sheet personas by gender), these approaches

are considerably more rudimentary than those offered when using digital data and systems to serve the personas to their users.

It is critical to keep an open mind when deciding on the chosen strategy for persona creation. Testing and experimenting with diverse methodologies is highly recommended for academics and practitioners, as breaking out of methodological comfort zones and adopting methods that are new to oneself is critical for learning and "becoming better" at creating personas over time and repetition. Even if accustomed to a particular method of persona creation, another method may be more relevant for the content and task at hand. Thus, researchers and practitioners may considerably gain from testing and exploring novel methods and reporting these experiences to others in the research community. To this end, we note that while it is straightforward to discover theoretical or conceptual treatises of personas' strengths and weaknesses, often tractable to a given methodological position, it is much more challenging to locate systematical empirical studies that would compare different persona creation methodologies within one study, using a fixed design task or context as a constant. Such studies are highly called for in order to make further progress in improving the persona creation methodologies.

Strategies and Opportunities for User Insights

To address **RQ3**, we discuss the opportunities of personas for user understanding for enhanced user insights (Shang & Chiu, 2022) for information systems.

Nylén and Holmström (2015) present a five-component framework concerning how firms can leverage user insights for digital innovation. This framework includes (1) *rich UX as measured by engagement, aesthetics, and usability*; (2) *value propositions for digital offerings involving value chain partners and customer segmentation*; (3) *intelligence on digital trends involving hardware, channels, technologies of interest, and user behaviors*; (4) *dynamic innovation teams that engage in unbroken learning*; and (5) *learning-by-doing*, while safeguarding resources for projects that demonstrate early signs of success.

These five components provide a path for personas to contribute value to digital innovation. This is because the components rely on an appropriate level of situational awareness concerning the needs of partners, customers, and the organization's talent. These stakeholder sets can be modeled via personas, offering multiple ways to support organizations' digital innovation efforts. Personas support UX design that is based on user requirements (Aoyama, 2007). Personas also help in formulating personified value propositions (Salminen, Kaate, et al., 2020), codifying digital user behaviors (Zhu et al., 2019), promoting collaborative user-centered design within digital innovation teams (Long, 2009).

Some of the promising avenues for employing personas for user insights include:

- **Augmenting the level of customer-centricity and market orientation** (Kohli & Jaworski, 1990) through the presentation of customer portraits that can evoke empathy.

- **Categorizing and segmenting student needs** in e-learning and remote education (Winkler & Söllner, 2018) to improve learning processes and outcomes.

- **Growing managers' immersion with user data** by making the personas responsive and interactive to end-user queries (Jansen et al., 2020), merging digital assistants with personas, and offering voice- and text-based user interfaces (Maedche et al., 2019).

- **Humanizing information technology artifacts** (Söllner et al., 2012) by giving a face to the dimensions of performance, process, and purpose. Thus, increasing trust in systems.

- **Personas for proficiency management and HR** – the high demand for IT professionals (Prommegger et al., 2020) sets pressure for defining functional and long-term matches between talent and organizations. Recently, the IT sector has witnessed a migration of different backgrounds, including various genders, cultures, and ethnicities. Personas can help profile and understand this ever-changing IT profession.

- **Representing stakeholder needs for requirements engineering** before starting the project and communicating these as personas that safeguard anonymity and facilitate dealing with sensitive topics. The more diverse the stakeholder groups are, the more useful personas can be (e.g., for global projects (Nielsen et al., 2013)) in reducing uncertainty, aligning expectations, and helping "name the pain" (Fernández et al., 2017).

- **Representing the user journey, product lifecycle, and service ecosystems** (Herterich et al., 2015) – especially the "dataification" of personas (Salminen, Guan, et al., 2020) enables modeling customer reactions in various stages of their journey, as well as understanding the motivations of service ecosystem and value networks.

- **Spanning knowledge gaps to serve users of information systems better** – for example, Schreieck et al. (2017) mention that refugees arriving in Europe struggle to obtain information from digital platforms, as these platforms are not necessarily adapted for their needs.

Generally, personas are one of several methods for bridging the division between IT systems and the users of these systems. The selection of a persona approach is driven by a mass of factors, as well as strengths and weaknesses of different (non-persona) approaches, organizational readiness, available resources, and attitudes within the organization for personas in general.

Future Research and Conclusion

Since their introduction, personas have been viewed as part of the digital innovation landscape, namely as an element of user-driven, user-centered design and innovation processes. To ensure methodological clarity, we examine three distinct approaches to persona creation: quantitative, qualitative, and mixed methodologies. We determine the strengths and weaknesses of each method by performing a comprehensive review of the available literature. We do a SWOT analysis of the three approaches in order to gain insight into the contexts in which each strategy should be used, and we present the results in a framework of guidelines for persona creators. As user populations become more complex and diverse, personas can be an invaluable tool for comprehending individuals in contexts of technology transformation.

The originality of our research contribution arises from the synthesis of HCI, MIS, and IS perspectives concerning understanding the value of personas for digital innovation. Future research should carry on this effort by deploying personas for digital innovation and system design projects. For example, are personas only beneficial for the early stage of the innovation process, or can personas provide value throughout the innovation cycle?

Future research on the methods' applicability could also examine application conditions, including aspects, for example, whether the personas are being used for NPD or improvement to an existing system. There are many other contexts to consider (Salminen et al., 2022). For instance, does it matter whether the project serves a large customer base, such as an e-

commerce company, or a smaller user base, such as a workout app for patients recuperating from a broken leg? Because the strategies' usefulness requires case-specific factors, there are no "one-size-fits-all" factors for persona creation. The main challenge, we surmise, is that the MIXED methodology is difficult to construe in practical applications at scale despite being theoretically superior to QUAL and QUANT methodologies.

On the one hand, it is relatively straightforward to carry out a qualitative analysis among the persona creation team because the team often shares a methodological background. On the other hand, it is relatively straightforward to build data science processes for QUANT personas, but it is very challenging to inject qualitative insights into this process. Thus, MIXED personas remain challenging. The successful implementation of MIXED personas requires following a couple of guiding principles, namely, (a) building cross-disciplinary persona creation teams that incorporate both strong qualitative and quantitative skills; (b) discovering new methods for researchers/designers to participate in co-creating personas with statistical algorithms and/or enriching data-driven personas with qualitative insights using interactive persona systems' user interfaces. There also remains an important need for experimental research that would dissect the effect of *starting with qualitative data vs. qualitative data*; that is, whether it is more fruitful to first explore qualitative or quantitative data in the persona creation process. With these conclusions, there remains much work for combining the strengths of the multiple persona creation methodologies in practical applications.

References

Adams, A., Lunt, P., & Cairns, P. (2008). A qualitative approach to HCI research. In P. E. Cairns & A. L. Cox (Eds.) *Research Methods for Human-Computer Interaction* (pp. 138-157). Cambridge University Press.

Aljohani, M., & Blustein, J. (2015). Personas help understand users' needs, goals and desires in an online institutional repository. *International Journal of Computer and Information Engineering, 9*(2), 629-636.

An, J., Kwak, H., Jung, S., Salminen, J., & Jansen, B. J. (2018). Customer segmentation using online platforms: Isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining, 8*(1), 1-19.

An, J., Kwak, H., Salminen, J., Jung, S., & Jansen, B. J. (2018). Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web (TWEB), 12*(4), 1-26.

Anvari, F., Richards, D., Hitchens, M., Babar, M. A., Tran, H. M. T., & Busch, P. (2017). An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software, 134*, 324-339.

Aoyama, M. (2007, October 15-19). *Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering.* Proceedings of the 15th IEEE International Requirements Engineering Conference (RE 2007)*,* Delhi, India.

Bamman, D., O'Connor, B., & Smith, N. A. (2013, August 4-9). *Learning Latent Personas of Film Characters.* Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria.

Baxter, K., Courage, C., & Caine, K. (2015). *Understanding Your Users: A Practical Guide to User Research Methods (2nd Edition)*. Morgan Kaufmann.

Blomquist, A., & Arvola, M. (2002, October 19). *Personas in action: Ethnography in an interaction design team.* Proceedings of the Second Nordic Conference on Human-Computer Interaction.

Brickey, J., Walczak, S., & Burgess, T. (2010, August 12-15). *A Comparative Analysis of Persona Clustering Methods.* Americas Conference on Information Systems (AMCIS2010), Lima, Peru.

Brickey, J., Walczak, S., & Burgess, T. (2012). Comparing Semi-Automated Clustering Methods for Persona Development. *IEEE Transactions on Software Engineering, 38*(3), 537-546.

Brooks, C., & Greer, J. (2014, March 24). *Explaining predictive models to learning specialists using personas.* Proceedings of the Fourth International Conference on Learning Analytics and Knowledge - LAK '14.

Chapman, C., Krontiris, K., & Webb, J. (2015). *Profile CBC: Using conjoint analysis for consumer profiles.* Sawtooth Software Conference Proceedings, Orlando, FL.

Chapman, C., Love, E., Milham, R. P., ElRif, P., & Alford, J. L. (2008). Quantitative evaluation of personas as information. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(16), 1107-1111.

Chapman, C., & Milham, R. P. (2006). The Personas' new clothes: Methodological and practical arguments against a popular method. In the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(5), 634-636.

Chu, E., Vijayaraghavan, P., & Roy, D. (2018). Learning personas from dialogue with attentive memory networks. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium.

21

Cleland-Huang, J., Czauderna, A., & Keenan, E. (2013). A persona-based approach for exploring architecturally significant requirements in agile projects. In J. Doerr & A. L. Opdahl (Eds.), *Requirements Engineering: Foundation for Software Quality* (pp. 18-33). Springer.

Cooper, A. (2004). *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity (2nd Edition)*. Sams Publishing.

Dang, D. P. T., Nguyen, L., Hoang, A. P., Pittayachawan, S., Akbari, M., & Nkhoma, M. (2021). Categorizing young Facebook users based on their differential preference of social media heuristics: A Q-Methodology approach. *Pacific Asia Journal of the Association for Information Systems, 13*(1), 71-96.

Dang-Pham, D., Pittayachawan, S., & Nkhoma, M. (2015). Demystifying online personas of Vietnamese young adults on Facebook: A Q-methodology approach. A*ustralasian Journal of Information Systems, 19*(1), 1-22.

Deng, S., Jiang, Y., Li, H., & Liu, Y. (2020). Who contributes what? Scrutinizing the activity data of 4.2 million Zhihu users via immersion scores. *Information Processing & Management, 57*(5), 1-14.

Dhakad, L., Das, M., Bhattacharyya, C., Datta, S., Kale, M., & Mehta, V. (2017, November 6). *SOPER: Discovering the Influence of Fashion and the Many Faces of User from Session Logs using Stick Breaking Process.* Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.

Drego, V. L., Dorsey, M., Burns, M., & Catino, S. (2010). *The ROI of Personas.* Forrester Research. https://www.forrester.com/report/The+ROI+Of+Personas/-/E-RES55359

Dupree, J. L., Devries, R., Berry, D. M., & Lank, E. (2016, May 7). *Privacy personas: Clustering users via attitudes and behaviors toward security practices.* Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.

Duvvuri, V. (2021). Expanding Digital Marketing Campaigns with Machine Learning Built Lookalike Audiences Having Varying Prior User Characteristics. In Á. Rocha, J. L. Reis, M. K. Peter, R. Cayolla, S. Loureiro, & Z. Bogdanović (Eds.), *Marketing and Smart Technologies (pp. 201-212).* Springer.

Fernández, D. M., Wagner, S., Kalinowski, M., Felderer, M., Mafra, P., Vetrò, A., Conte, T., Christiansson, M.-T., Greer, D., Lassenius, C., Männistö, T., Nayabi, M., Oivo, M., Penzenstadler, B., Pfahl, D., Prikladnicki, R., Ruhe, G., Schekelmann, A., Sen, S., … Wieringa, R. (2017). Naming the pain in requirements engineering: Contemporary problems, causes, and effects in practice. *Empirical Software Engineering, 22*(5), 2298-2338.

Floyd, I. R., Jones, C. M., & Twidale, M. B. (2008). Resolving incommensurable debates: A preliminary identification of persona kinds, attributes, and characteristics. *Artifact: Journal of Design Practice, 2*(1), 12-26.

Ford, D., Zimmermann, T., Bird, C., & Nagappan, N. (2017, November 9-10). *Characterizing software engineering work with personas based on knowledge worker actions.* Proceedings of the 11th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, Toronto, ON, Canada.

Gilsing, R., Turetken, O., Ozkan, B., Grefen, P., Adali, O. E., Wilbik, A., & Berkers, F. (2021). Evaluating the design of service-dominant business models: A qualitative method. *Pacific Asia Journal of the Association for Information Systems, 13*(1), 36-70.

Glaser, B. G., & Strauss, A. L. (2009). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Transaction Publishers.

Goodman-Deane, J., Waller, S., Demin, D., González-de-Heredia, A., Bradley, M., Clarkson, J. P., & Clarkson, J. P. (2018, June 25-28). Evaluating inclusivity using quantitative personas. *Proceedings of Design Research Society Conference 2018 , Limerick, Ireland.*

Goodwin, K., & Cooper, A. (2009). *Designing for the Digital Age: How to Create Human-Centered Products and Services*. John Wiley & Sons.

Grudin, J. (2006). Why Personas Work: The Psychological Evidence. In J. Pruitt & T. Adlin (Eds.), *The Persona Lifecycle (pp. 642-663).* Elsevier.

Guo, H., & Razikin, K. B. (2015, December 7). *Anthropological user research: A data-driven approach to personas development.* Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction.

Guo, J., & Yan, P. (2011, May 11-13). *User-centered information architecture of university library website.* 2011 3rd International Conference on Computer Research and Development, Shanghai, China.

Hassandoust, F., Techatassanasoontorn, A., & Tan, F. B. (2016). Factors Influencing the infusion of information systems: A literature review. *Pacific Asia Journal of the Association for Information Systems, 8*(1), 1-32.

Hazrati, N., & Ricci, F. (2022). Recommender systems effect on the evolution of users' choices distribution. *Information Processing & Management, 59*(1), 1-18.

Herterich, M. M., Uebernickel, F., & Brenner, W. (2015). The Impact of Cyber-physical Systems on Industrial Services in Manufacturing. *Procedia CIRP, 30*, 323-328.

Hill, C. G., Haag, M., Oleson, A., Mendez, C., Marsden, N., Sarma, A., & Burnett, M. (2017, May 7). *Gender-inclusiveness personas vs. stereotyping: Can we have it both ways?* Proceedings of the 2017 CHI Conference on human factors in computing systems.

Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization, 28*(1), 52-61.

Hirskyj-Douglas, I., Read, J. C., & Horton, M. (2017, July 3-6). *Animal personas: Representing dog stakeholders in interaction design.* Proceedings of the 31st British Computer Society Human Computer Interaction Conference, Sunderland, UK.

Hisham, S. (2009, November 23). *Experimenting with the use of persona in a focus group discussion with older adults in Malaysia.* Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design: Open 24/7.

Ho, S.-C., & Hsu, Y.-P. (2022). Paving the way for digital transformation: Investigate Customer experiences of using mobile apps. *Pacific Asia Journal of the Association for Information Systems, 14*(1), 18-39.

Holden, R. J., Kulanthaivel, A., Purkayastha, S., Goggins, K. M., & Kripalani, S. (2017). Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure. *International Journal of Medical Informatics, 108*(December), 158-167.

Holmgard, C., Liapis, A., Togelius, J., & Yannakakis, G. N. (2014, August 26-29). *Evolving personas for player decision modeling.* 2014 IEEE Conference on Computational Intelligence and Games (CIG), Dortmund, Germany.

Holtzblatt, K., Beringer, J., & Baker, L. (2005, April). *Rapid user centered design techniques: Challenges and solutions.* CHI'05 Extended Abstracts on Human Factors in Computing Systems.

Huh, J., Kwon, B. C., Kim, S.-H., Lee, S., Choo, J., Kim, J., Choi, M.-J., & Yi, J. S. (2016). Personas in online health communities. *Journal of Biomedical Informatics, 63*(October), 212-225.

Idoughi, D., Seffah, A., & Kolski, C. (2012). Adding user experience into the interactive service design loop: A persona-based approach. *Behaviour & Information Technology, 31*(3), 287-303.

Ishii, R., Ito, S., Ishihara, M., Harada, T., & Thawonmas, R. (2018, August 14-17). *Monte-Carlo Tree Search Implementation of Fighting Game AIs Having Personas. 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, Netherlands.*

Jansen, B. J., Jung, S., Salminen, J., Guan, K., & Nielsen, L. (2021, January 5-8). *Strengths and weaknesses of persona creation methods: Outlining guidelines for novice and experienced users and opportunities for digital innovations.* Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021).

Jansen, B. J., Salminen, J., & Jung, S. (2020). Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics. *Data and Information Management, 4*(1), 1-17.

Kim, E., Yoon, J., Kwon, J., Liaw, T., & Agogino, A. M. (2019). From innocent irene to parental patrick: Framing user characteristics and personas to design for cybersecurity. *Proceedings of the Design Society: International Conference on Engineering Design, 1*(1),1773-1782.

Kohli, A. K., & Jaworski, B. J. (1990). Market orientation: The construct, research propositions, and managerial implications. *The Journal of Marketing, 54*(2), 1-18.

LeRouge, C., Ma, J., Sneha, S., & Tolle, K. (2013). User profiles and personas in the design and development of consumer health technologies. *International Journal of Medical Informatics, 82*(11), e251-e268.

Li, H., Chen, Q., Zhong, Z., Gong, R., & Han, G. (2022). E-word of mouth sentiment analysis for user behavior studies. *Information Processing & Management, 59*(1), 1-12.

Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., & Dolan, B. (2016). *A persona-based neural conversation model.* Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.

Li, Y., Yuan, X., & Che, R. (2021). An investigation of task characteristics and users' evaluation of interaction design in different online health information systems. *Information Processing & Management, 58*(3), 1-20.

Long, F. (2009). *Real or imaginary: The effectiveness of using personas in product design.* Proceedings of the Irish Ergonomics Society Annual Conference, Dublin, Ireland.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering, 61*(4), 535-544.

Matthews, T., Judge, T., & Whittaker, S. (2012, May 5). *How do designers and user experience professionals actually perceive and use personas?* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

McGinn, J. J., & Kotamraju, N. (2008, April 6). *Data-driven persona development.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Mesgari, M., Okoli, C., & Guinea, A. O. de. (2015, August 13-15). *Affordance-based User Personas: A mixed-method approach to persona development.* Twenty-first Americas Conference on Information Systems, Puerto Rico.

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies, 32*(5), 417-430.

Miaskiewicz, T., & Luxmoore, C. (2017). The use of data-driven personas to facilitate organizational adoption–A case study. *The Design Journal, 20*(3), 357-374.

Miaskiewicz, T., Sumner, T., & Kozar, K. A. (2008, April 6). *A latent semantic analysis methodology for the identification and creation of personas.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Mijač, T., Jadrić, M., & Ćukušić, M. (2018, May 21-25). *The potential and issues in data-driven development of web personas.* 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia.

Minichiello, A., Hood, J. R., & Harkness, D. S. (2018). Bringing user experience design to bear on STEM education: A narrative literature review. *Journal for STEM Education Research, 1*(1), 7-33.

Mulder, S., & Yaar, Z. (2006). *The User is Always Right: A Practical Guide to Creating and Using Personas for the Web.* New Riders.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *MIS Quarterly, 41*(1), 223-238.

Neate, T., Bourazeri, A., Roper, A., Stumpf, S., & Wilson, S. (2019, May). *Co-created personas: Engaging and empowering users with diverse needs within the design process.* Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Ng, C. S. P., & Wang, E. T. (2019). Impact of social media management styles on willingness to be a fan: A transaction cost economics perspective. *Pacific Asia Journal of the Association for Information Systems, 11*(2), 6-38.

Nielsen, L. (2004). Engaging personas and narrative scenarios [PhD Thesis, Samfundslitteratur]. http://personas.dk/wp-content/samlet-udgave-til-load.pdf

Nielsen, L. (2013). *Personas—User Focused Design*. Springer-Verlag.

Nielsen, L., Nielsen, K. S., Stage, J., & Billestrup, J. (2013). Going Global with Personas. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013* (Vol. 8120, pp. 350-357). Springer, Berlin, Heidelberg.

Nylén, D., & Holmström, J. (2015). Digital innovation strategy: A framework for diagnosing and improving digital product and service innovation. *Business Horizons, 58*(1), 57-67.

Poulain, R., & Tarissan, F. (2020). Investigating the lack of diversity in user behavior: The case of musical content on online platforms. *Information Processing & Management, 57*(2), 1-18.

Prommegger, B., Wendrich, M., Wiesche, M., & Krcmar, H. (2020, June 19). *Short-term Affair or Long-term Commitment? An Investigation of Employees without IT Background in IT Jobs.* Proceedings of the 2020 on Computers and People Research Conference.

Pruitt, J., & Adlin, T. (2010). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design.* Morgan Kaufmann.

Pruitt, J., & Grudin, J. (2003, June). *Personas: Practice and theory.* Proceedings of the 2003 Conference on Designing for User Experiences.

Rahimi, M., & Cleland-Huang, J. (2014, September 15). *Personas in the middle: Automated support for creating personas as focal points in feature gathering forums.* Proceedings of the 29th ACM/IEEE International Conference on Automated Software Engineering.

Salminen, J., Guan, K., Jung, S., Chowdhury, S. A., & Jansen, B. J. (2020, April). *A literature review of quantitative persona creation.* CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.

Salminen, J., Guan, K. W., Jung, S.-G., & Jansen, B. J. (2022). *Use cases for design personas: A systematic review and new frontiers.* 2022 ACM Conference on Human Factors in Computing Systems (CHI'22), New Orleans, USA.

Salminen, J., Jung, S., Chowdhury, S. A., Sengün, S., & Jansen, B. J. (2020, April 23). *Personas and analytics: A comparative user study of efficiency and effectiveness for a user identification task.* Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.

Salminen, J., Jung, S., & Jansen, B. J. (2019, January). *The future of data-driven personas: A marriage of online analytics numbers and human attributes.* Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019).

25

Salminen, J., Kaate, I., Kamel, A. M. S., Jung, S., & Jansen, B. J. (2020). How does personification impact ad performance and empathy? An experiment with online advertising. *International Journal of Human–Computer Interaction, 37*(2), 141-155.

Salminen, J., Şengün, S., Kwak, H., Jansen, B. J., An, J., Jung, S., Vieweg, S., & Harrell, D. F. (2018). From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *First Monday, 23*(6).

Schreieck, M., Wiesche, M., & Krcmar, H. (2017). Governing nonprofit platform ecosystems – an information platform for refugees. *Information Technology for Development, 23*(3), 618-643.

Seiden, J., & Gothelf, J. (2003). *Lean UX: Applying Lean Principles to Improve User Experience.* Oreilly.

Shang, S. S. C., & Chiu, L. S. L. (2022). *Leveraging Smart Technology for User Experience Personalization – A Comparative Case Study of Innovative Payment Systems. Pacific Asia Journal of the Association for Information Systems, 14*(1), 105-125.

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. (2012). *Understanding the Formation of Trust in IT Artifacts.* Thirty Third International Conference on Information Systems, Orlando, FL.

Sun, X., Zhou, X., Wang, Q., Sharples, S., & Zhang, Y. (2021). A review of research methodologies employed in serendipity studies in the context of information research. *Pacific Asia Journal of the Association for Information Systems, 13*(2), 72-90.

Tan, H., Peng, S., Liu, J.-X., Zhu, C.-P., & Zhou, F. (2021). Generating Personas for Products on Social Media: A Mixed Method to Analyze Online Users. I*nternational Journal of Human–Computer Interaction,* 1-12.

Tashakkori, A., & Teddie, C. (1998). *Mixed Methodology.* Sage Publication.

Tempelman-Kluit, N., & Pearce, A. (2014). Invoking the User from Data to Design. *College & Research Libraries, 75*(5), 616-640.

Thoma, V., & Williams, B. (2009). Developing and Validating Personas in e-Commerce: A Heuristic Approach. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2009 (pp. 524–527)*. Springer, Berlin Heidelberg.

Tu, N., He, Q., Zhang, T., Zhang, H., Li, Y., Xu, H., & Xiang, Y. (2010, November 26-28). *Combine qualitative and quantitative methods to create persona.* 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, Kunming, China.

Tychsen, A., & Canossa, A. (2008, November 3). *Defining personas in games using metrics.* Proceedings of the 2008 Conference on Future Play: Research, Play, Share.

Vallet, F., Khouadjia, M., Amrani, A., & Pouzet, J. (2021). Designing a data visualization and analysis tool for supporting decision making with public transportation network. *Proceedings of the Design Society, 1*, 1093-1102.

Vecchio, P. D., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management, 54*(5), 847-860.

Vosbergen, S., Mulder-Wiggers, J. M. R., Lacroix, J. P., Kemps, H. M. C., Kraaijenhagen, R. A., Jaspers, M. W. M., & Peek, N. (2015). Using personas to tailor educational messages to the preferences of coronary heart disease patients. *Journal of Biomedical Informatics, 53*, 100-112.

Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., & Foster, J. (2021). Clustering and classifying users from the national museums liverpool website. In G. Berget, M. M. Hall, D. Brenn, & S. Kumpulainen (Eds.), *Linking Theory and Practice of Digital Libraries (pp. 202-214)*. Springer International Publishing.

Wang, L., Li, L., Cai, H., Xu, L., Xu, B., & Jiang, L. (2018, October 1-2). *Analysis of regional group health persona based on image recognition.* 2018 Sixth International Conference on Enterprise Systems (ES), Limassol, Cyprus.

Watanabe, Y., Washizaki, H., Honda, K., Noyori, Y., Fukazawa, Y., Morizuki, A., Shibata, H., Ogawa, K., Ishigaki, M., Shiizaki, S., Yamaguchi, T., & Yagi, T. (2017, May 23). *ID3P: Iterative data-driven development of persona based on quantitative evaluation and revision.* Proceedings of the 10th International Workshop on Cooperative and Human Aspects of Software Engineering, Buenos Aires, Argentina.

Winkler, R., & Söllner, M. (2018). *Unleashing the potential of chatbots in education: A State-Of-The-Art analysis.* Academy of Management Annual Meeting (AOM), Chicago, USA.

Wöckl, B., Yildizoglu, U., Buber, I., Aparicio Diaz, B., Kruijff, E., & Tscheligi, M. (2012, October). *Basic senior personas: A representative design tool covering the spectrum of European older adults.* Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility.

Wright, P., & McCarthy, J. (2008, April). *Empathy and experience in HCI.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Xie, I., Babu, R., Lee, T. H., Castillo, M. D., You, S., & Hanlon, A. M. (2020). Enhancing usability of digital libraries: Designing help features to support blind and visually impaired users. *Information Processing & Management, 57*(3), 1-14.

Xtensio. (2021). *How to Create a User Persona.* Xtensio. https://xtensio.com/how-to-create-a-persona/

Zaugg, H., & Ziegenfuss, D. H. (2018). Comparison of personas between two academic libraries. *Performance Measurement and Metrics, 19*(3), 142-152.

Zhang, X., Brown, H.-F., & Shankar, A. (2016, May). *Data-driven personas: Constructing archetypal users with clickstreams and user telemetry.* Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA.

Zhu, H., Wang, H., & Carroll, J. M. (2019, June). *Creating persona skeletons from imbalanced datasets— A case study using U.S. older adults' health data. Proceedings of the 2019 on Designing Interactive Systems Conference, San Diego, CA, USA.*

About the Authors

Mr. Bernard J. Jansen is a Principal Scientist in the social computing group of the Qatar Computing Research Institute. He is a graduate of West Point and has a Ph.D. in computer science from Texas A&M University. Professor Jansen is editor-in-chief of the journal Information Processing & Management (Elsevier).

Mr. Soon-gyo Jung is a software engineer focused on data-driven/data-intensive systems in the social computing group at Qatar Computing Research Institute (QCRI), Doha, Qatar. He is currently strongly interested in computational social science, especially exploring its significant impact on society and how people communicate and share their culture with others. He received a B.E. degree in computer software from the Kwangwoon University, Seoul, Korea, in 2014, and a M.S. degree in electrical and computer engineering from the Sungkyunkwan University, Suwon, Korea, in 2016.

Ms. Kathleen Guan is currently, as of the date of this article, a graduate student with the University College London. She has a Bachelor of Science in Foreign Service in International Law from Georgetown University, and research training in Public Health from Johns Hopkins University. In addition to research consulting for industry, Kathleen is currently a research student in Neuroscience and Psychopathology through a joint graduate program between University College London and Yale School of Medicine.

Dr. Lene Nielsen is an Associate Professor of Business IT and Technology Innovation Management and Entrepreneurship at the IT University of Copenhagen. She has been researching personas for years and is Denmark's foremost expert on persona design.

Dr. Joni Salminen is a faculty member of the School of Marketing and Communication, University of Vaasa, Finland, and he is also affiliated with the Turku School of Economics at the University of Turku. Previously, he was a scientist at the Qatar Computing Research Institute. His research interests relate to personas, human-computer interaction, online hate, user and customer segmentation, and social media platforms.