WHY ARE PERSONAS THE WAY THEY ARE? IDENTIFYING SIX PERSONA CREATION STRATEGIES

Essi Häyhänen University of Vaasa, Joni Salminen University of Vaasa

Bernard J. Jansen Qatar Computing Research Institute, Hamad Bin Khalifa University

ABSTRACT

User persona creation is a well-established technique in user-centred design. However, the persona creation process is not well understood. Addressing this gap, this study investigates how twenty-nine student designers created personas from a customer dataset. Six main persona creation strategies emerged: (1) data-oriented, (2) diversity, (3) imaginative, (4) sociability, (5) self-centric, and (6) mixed strategy. The most common was the mixed strategy that combined data-oriented and diversity strategies to create personas based on the presented data while representing demographically diverse user groups. Most commonly, the designers created four personas, and they aimed at symmetry and evenness by creating, for example, an even number of male and female personas. The results demonstrate that creator-provided reasonings can increase the transparency of the persona-creation process, which could support stakeholder acceptance of the personas.

KEY WORDS

User Personas, Persona Creation, Strategies, User-Centred Design

INTRODUCTION

Persona creation is a prominent user-centred design (UCD) technique in software development, human-computer interaction (HCI), marketing, and other domains (Cooper 1999; Märtin, Bissinger & Asta 2021; Siricharoen 2021). Personas are attributed with several benefits, including increasing stakeholder (e.g., designer, developer, marketer) empathy toward the target user group that the persona represents (Melo et al. 2020; Hiléia da Silva Melo et al. 2021; Nolte et al. 2022), making user information more communicable (Karahasanović, Følstad & Schittekat 2021; MacDonald, Rose & Putnam 2022), mitigating stakeholder self-referential bias when making design choices (Curry, Robertson & Rieser 2020; Emmanuel & Polito 2022), and creating a shared mental model (Moore, Barbour & Lee 2017) to keep users or customers in mind when none are physically available to give feedback (Damiano et al. 2022; Villareale, Harteveld & Zhu 2022).

Persona development involves creating fictitious people that represent the different types of people for whom a product or service is intended. Persona creation has received much attention from scholars (Mulder & Yaar 2006; Nielsen & Storgaard Hansen 2014; Nielsen 2019;

B. J. Jansen et al. 2021), forming a central focal point for persona studies (Marshall, Moore & Barbour 2019).

The theoretical foundation of persona creation draws upon interdisciplinary perspectives from HCI, marketing, and design, leaving other areas for future research. The work focuses on user personas which is a recognised subset of the fictitious persona typology within persona studies (Giles 2020). We clarify that, in addition to user personas, a major fraction of 'persona studies' investigates how individuals construct online personas (Marshall et al. 2019). The key distinction between "online personas" and "user personas" is that the former deals with the representation of self, whereas the latter deals with the representation of others; more specifically, other groups of people. In HCI, persona creation is a widely used technique in UCD to foster stakeholder empathy and communication (Cooper 1999; Nielsen 2019). For example, a designer creating a healthcare app might develop "Maria, 65, tech-hesitant retiree with chronic arthritis" as a primary persona to ensure the interface accommodates users with limited mobility and digital experience.

However, the individual-level persona creation process remains underexplored, particularly regarding the persona creators' subjective interpretations and the role of transparency in persona development (B. Jansen et al. 2021). Design disciplines emphasize the importance of personas as tools for decision-making, ensuring that design processes reflect the real needs of users while mitigating cognitive biases (Hiléia da Silva Melo et al. 2021).

The broad literature on persona creation (Cooper 1999; Mulder & Yaar 2006; Nielsen 2019; Salminen, Guan, et al. 2020; B. J. Jansen et al. 2021; Jansen et al. 2022) shows that most studies focus on two aspects: (a) the comparison of different persona creation methods in terms of their strengths, weaknesses, and applicability and (b) proposing steps or procedures for an "ideal" persona creation process (Salminen, Guan, et al. 2020; Salminen et al. 2021). However, less attention has been given to understanding how and why creators choose certain information for the personas, how they choose the number of personas, and what reasonings this process entails. These questions are decisive, especially in manual persona creation (Nielsen 2019), which refers to the creation of personas without the use of an algorithmic method and which remains the dominant form of persona creation (Hiang, Kulathuramaiyer & Zaman 2017). Thus, *how do different people create personas from the same user data?* This question motivates the current study, leading to two research questions (RQs):

- *RQ1: What types of personas do different student creators create from the same data?*

- *RQ2: What kind of strategies do student creators apply in their persona creation process?*

To address the RQs, we conducted a study in which 29 university students created personas from a customer dataset. We then compare the persona attributes and ask the persona creators to explicate their choices for the persona attributes. Our study provides primary information on this understudied topic of the reasonings people gave to their persona creation choices. The findings inform persona researchers of the intricacies beyond outlining generic "steps" for the persona creation. Various authors have proposed steps in the persona creation process but *how* people actually follow them is not understood. Therefore, we believe that addressing these RQs is of value to the persona studies research community.

LITERATURE REVIEW

Prior research highlights the range of approaches to persona creation, from qualitative techniques such as ethnography and grounded theory (White & Devitt 2021) to quantitative methods that use statistical analysis and clustering algorithms (An, Kwak, Salminen, et al. 2018; Salminen, Rao, et al. 2020). While these approaches provide structured frameworks, the actual process of persona creation is often influenced by the creators' interpretation of the data, their personal values, and contextual factors (Bødker et al. 2012; Nielsen 2019). Furthermore, literature has emphasized the non-linear and sometimes idiosyncratic nature of persona creation (Chapman & Feit 2019), which emphasizes the importance of understanding the reasoning behind persona design choices. As proposed in the literature, the distinction between the ideal persona creation processes and the more flexible, subjective approaches observed in practice (Jansen et al. 2022) forms a critical gap that this study aims to address.

While the 'persona creation method' deals with the overall approach to the persona creation (B. J. Jansen et al. 2021), the 'persona creation process' refers to the steps taken within that method from the start to obtaining a ready set of personas. Instead of one 'go-to' methodology, research studies have used a variety of processes (Long 2009; Bødker et al. 2012; Nielsen & Storgaard Hansen 2014; Seidelin et al. 2014; Tarkkanen et al. 2018). However, these processes share similarities. These processes are often presented as rigorous, systematic steps (typically in the range of 4-6 steps). Most include activities such as identifying target consumers, collecting user data, categorizing users into personas, and creating and presenting persona details (Brickey, Walczak & Burgess 2012). In the following, we describe a standard or "ideal" five-step process as often displayed in literature (e.g., Pruitt & Adlin 2006).

The first step typically is to determine the purpose and scope of the persona creation process, as well as defining the target audience and the context of using personas (Jansen et al. 2022). Establishing the purpose of the personas helps designers pinpoint their objectives, the range of the personas they wish to develop, and the appropriate data sources (Pruitt & Adlin 2010).

The second step typically involves data collection. In the case of qualitative persona creation, data is usually collected manually through interviews, surveys with small sample sizes, and focus groups. For quantitative persona creation, the most common method is a survey (Salminen, Guan, et al. 2020). The third approach is a mixed method approach encompassing quantitative and qualitative data. In a typical mixed-method approach, quantitative data collection is followed by data enrichment (contextualization) through qualitative means (Jansen et al. 2022).

The third step typically involves data analysis. Qualitative approaches for persona creation may include grounded theory (White & Devitt 2021), ethnography, and narrative analysis (Rapp 2019), among others. In turn, quantitative data is analysed through statistical or computational approaches, e.g., K-means clustering, hierarchical clustering, and principal component analysis (Salminen, Guan, et al. 2020). As the name suggests, the mixed methods approach employs both qualitative and quantitative analysis techniques in a joint effort (Jansen et al. 2022).

The fourth step typically involves developing rough segments. These segments are based on the data type used in the process. In the case of qualitative persona creation, segments are often developed through manual analysis, with their number based on the creator's judgment. In quantitative persona creation, creators often aim at a numerically optimal number

of segments. Since the mixed-method approach also involves quantitative data, it can result in a specific number of segments (B. J. Jansen et al. 2021).

The fifth and final step typically involves the persona write-up, also known as enrichment. The developed segments are further enhanced by giving them names, pictures, behavioural qualities, and other characteristics that lead to the creation of rounded persona profiles (Nielsen 2019). This enrichment gives the personas their final form (Salminen et al. 2018). This step can be very similar regardless of applying the qualitative, quantitative, and mixed-method approaches.

We present central caveats about this typical process described in the literature. First, though the above characterization tends to apply to many studies, the steps to create personas may vary depending on the approach and data type. Second, the steps are typically followed by (1) persona validation, i.e., evaluation of the created personas' quality and (2) persona implementation, i.e., putting the personas in use. Third, and most importantly, though most studies present persona creation as a systematic, logical, organized, and disciplined process, in reality, the persona creation process is often 'murky', non-linear, and varied with many subjectivities (i.e., influenced by personal feelings, tastes, biases, or opinions) and human factors like values, interpretations, emphases, and abilities of the individuals involved in the persona creation effort.

Persona creators are, to an extent, fallible humans subjected to information about users from a variety of sources, including self-observation, friends, technology media, marketing organizations, analyst reports, conferences, and support cases (Chapman & Milham 2006). These diverse inputs shape their impressions of users, leading to idiosyncratic creation processes and adverse outcomes such as stereotypical and biased personas (Turner & Turner 2011). Yet, the though-processes of the individuals creating the personas are not well-examined in the literature, so we lack information about the true of the thinking behind crucial design choices that shape the set of created personas. Understanding these factors would significantly advance the current persona design theory. Previous research in persona creation typically remains at the level of comparing the different persona creation approaches–the methods themselves–or reporting and documenting the steps and their outcomes in a systematic and logical manner: they describe the process from the outside, not aiming at shedding light to the persona creators' thinking process and reasoning of choices involved when creating a set of personas. In the current study, we focus on the creators' reasoning behind the persona creation process to address this gap. In so doing, we address the lack of empirical research on how persona creators process and use information about users when creating personas and how this information is integrated into the persona creation process; in other words, *what thought processes do persona creators follow?*

METHOD

Dataset for Persona Creation

The dataset was sourced from *Kaggle*, a well-known machine learning and data science community, where it was made publicly available for any number of purposes, including research. The dataset is based on fictitious customer records (n = 10,695). It is among the most popular customer segmentation datasets in *Kaggle*, implying that it is commonly used and represents a standard or customary customer segmentation dataset. (As a side note, we note that publicly available datasets for persona creation are scarce, as identified in related work (Salminen et al. 2021); the use of a customer segmentation dataset for a persona creation task appears justifiable given the similarities that customer segmentation and persona creation

share, i.e., both dealing with grouping people data in a way that serves decision making about said people.) As can be seen from Table 1, the dataset contains variables that (a) are commonly used for customer segmentation (both demographic and behavioural) and (b) represent different data types (integers, binary classes, multiple classes), making the dataset exemplary of a typical customer segmentation dataset.

Table 1: Description of the fictitious customer dataset. The variables include typical demographic and sociographic customer information.

Variable	Description	Possible values
ID	Customer Identification	Numerical
Gender	Customer's gender identity	Male/Female
Ever Married	Whether the customer has been married	Yes/No
Age	Customer's age	Integer
Graduated	Whether the customer has graduated from university	Yes/No
Profession	Customer's profession	Healthcare, Engineer, Lawyer, Entertainment, Artist, Executive, Doctor, Homemaker, Marketing
Work Experience (years)	Customer's work experience	Integer
Spending Score	Customer's amount of spending in the fictitious company	Low/Average/High
Family Size	Number of people in the customer's family	Integer

Participants

Twenty-nine (n=29) students participated in this study as part of their normal business course assignment and agreed to use it for research. The students included as well as exchange students from different nationalities, mostly European (e.g., France, Spain). There was a roughly equal representation of male and female participants in their 20s. The participants were bachelor's degree students taking an elective digital marketing course. Their majors included marketing, communication, and management. The students were advised that their participation could be used for research purposes ("The data collected based on student participation may be used as part of ongoing research on data-driven personas. In this case, the data will be treated anonymously, and no personally identifiable information will be disclosed."). Appendix 1 (https://osf.io/vuqy4/?view_only=eda9e83c275b4fbbb9ab6094679da5f6) shows the scenario provided to the students.

In addition to these instructions, an example persona layout was provided to the students based on a common template for designing personas (Nielsen et al. 2015). Each student worked independently to create the personas and had approximately two weeks to complete the task. The students returned the persona profiles they had created along with a spreadsheet information of the personas. In addition, they were asked to answer a survey in which they could explain their persona creation process in detail. The persona assignment constituted 20% of the course grading, so we believe the students carried out the task carefully and gave deliberate thought to how to create personas from the source data. Examples of the created personas are shown in Figures 1-4.

High score PERSONA

	Patrick
Gender **Age** **Work** **Marital status** **Home**	• Male • 45 years old • Executive, 2-year experience • Married, two children of 12 and 5 years old • FRANCE, Paris
Biography	Patrick works as a sales director in a French company. Married and father of 2 children. He lives in the suburbs of Paris. He likes sports and music. He likes his job very much. He hopes to be able to find some time to do more sport and go more often to the gym
Motivations and expectations	• Wants to keep in shape, and go to the gym at least 4 time per week • Need a coach that can help him to understand his body and be more efficient
Barriers and reluctance	• His own motivation is the first barrier • Finds few coaches for experienced sportsmen • Lack of advice and personalised exercises to perform a sport
Personality	Social, bon vivant and pragmatic Practices yoga

Figure 1: Persona profile examples created by student designer P03

Figure 2: Persona profile examples created by student designer P06

Figure 3: Persona profile examples created by student designer P17

Figure 4: Persona profile examples created by student designer P21

RESULTS

RQ1: What type of personas do different student creators create from the same data?

In total, the students created 120 personas. On average, the students created 4.14 personas (SD=0.90). The most common number of personas created (i.e., mode) was four. The maximum was six personas, and the minimum was two personas. Figure 5 shows the distribution of the personas created. Of the 29 persona sets created by the student designers, the overwhelming majority (n=21, 72.4%) contained an even number of personas (i.e., 2, 4, 6), while the persona sets that contained an odd number of personas (i.e., 3, 5) were much rarer (n=8, 27.6%).

Figure 5: The number of personas created by the students. Four personas were the most common number.

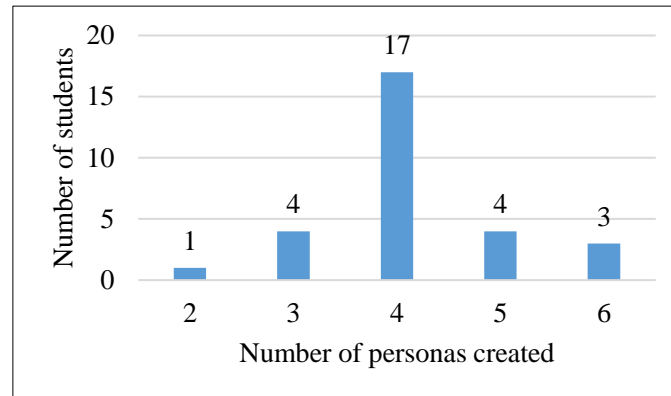The average age of personas was 40.69 years (SD = 13.80). The RDS (relative standard deviation) was 0.34 among all the personas, which is relatively high, implying that the personas varied substantially by age. The youngest persona created was 19 years old, while the oldest was 86 years old. One participant (P27) did not include numerical age values but instead used age ranges (middle-aged, 60+, below 35, and 25-55). We used mean-based imputation to replace these values with the average age of the personas (M = 41). A one-way analysis of variance (ANOVA) test results indicates that there was no significant difference in the average ages of personas by different participants, $F_{(28,91)} = 0.50$, $p = 0.98$.

Of all the personas created, 63 (52.50%) were male, and 57 (47.50%) were female. All participants included at least one of the two available genders (male or female) in their personas.

There were 19 different nationalities among the created personas: the US, France, Spain, UK, Finland, Australia, Germany, Sweden, Turkey, Poland, Italy, Scotland, Mexico, Belgium, Czech Republic, Nigeria, Nepal, Canada, and Switzerland.

Of the personas created, 62 (51.67%) were married and 58 (48.33%) were not married. Regarding graduation status, 91 (75.83%) were graduates and 29 (24.17%) were non-graduated. The average work experience of the personas was 5.82 years (SD = 8.22 years), and the average family size was 2.33 (SD = 1.20). In terms of the spending score, 59 (49.20%) personas were in the Low segment, 7 (5.80%) in the Average segment, and 54 (45.0%) in the High segment. The most popular profession was Artist (n = 25, 20.83%), while the least popular profession was Homemaker which was not selected even once. In addition to the professions given to the students, some chose professions outside this list, including Student (n = 5),

Architect (n = 1), Retired banker (n = 1), Painter (n = 1), Barman (n = 1), Self-employed creator of a clothing brand (n = 1), Worker (n = 1), and Teacher (n = 1).

RQ2: What kind of strategies do student creators apply in their persona creation process?

We analysed the open-ended responses left by the students in the questionnaire that ask how the student determined each persona attribute (e.g., "How did you determine the personas' age?"). Based on our analysis, six main persona-creation strategies emerged (see Table 2). The creation of the strategies adopted the guidelines of qualitative coding by Glaser and Strauss (1967), specifically in coding the material, grouping the codes, naming the groups, and obtaining saturation (i.e., new codes would not emerge). The lead author analysed the reasonings by the participants, and provided the strategies, definitions, and examples to other authors.

Table 2: Persona creation strategies and sub-strategies

Strategy	Definition	Sub-strategy	Definition
S01: Data-oriented strategy	Using quantitative data (e.g., frequencies, proportions, descriptive statistics) as the starting point for persona creation	S01a: Strategy of central tendency	Using mean, mode, median, and other possible indicators of what is 'typical' or 'normal' in the dataset to generate personas that represent this normality
		S01b: Secondary data strategy	Using data, information, or statistics not directly related to the primary data to justify their design choices
		S01c: Data associations	Using previously determined persona information on one attribute as a source of inspiration for another attribute
S02: Diversity strategy	Creating personas that lack major demographic biases.	S02a: Gender balance	Creating persona sets with an equal representation of all available genders
		S02b: Age diversification	Creating persona sets with a representation of all available age ranges (or at least multiple ages)
		S02c: Ethnic and cultural diversity	Creating persona sets with a representation of different races, ethnic backgrounds, or nationalities
		S02d Comparative strategy	Creating personas so that they can be compared to bring out the differences among the segments in the dataset
S03: Imaginative strategy	Using fictitious (made-up) data or content as persona information	S03a: Random selection strategy	Choosing the persona information at random
		S03b: Common sense strategy	Relying on what is typical for 'people of this type'; i.e., general knowledge about people

Strategy	Definition	Sub-strategy	Definition
S04: Sociability strategy	Attempting to make the persona socially acceptable	N/A	
S05: Self-centric strategy	Relying on personal experiences to determine the persona information	S05a: Stereotyping strategy	Relying on typical information about a certain group of people
		S05b: Realistic strategy	Creating as realistic personas as possible ("real-world personas") based on the creator's experiences
S06: Mixed strategy	Combining elements from different strategies, e.g., 'data-oriented strategy' and 'imaginative strategy'	N/A	

We named these strategies (1) data-oriented strategy, (2) diversity strategy, (3) imaginative strategy, (4) sociability strategy, (5) self-centric strategy, and (6) mixed strategy. Students in the first category created personas by using quantitative data. This strategy can be divided into three sub-strategies: (a) strategy of central tendency, (b) secondary data strategy, and (c) data associations. Students in the second category (secondary data strategy) created personas by considering diversity, e.g., gender balance, age diversification, or ethnic and cultural diversity.

Students in the third category created personas by utilizing fictitious (made-up) data or content as persona information. This imaginative strategy can be divided into two sub-strategies: (a) random selection strategy and (b) common sense strategy.

Students in the fourth category attempted to create socially acceptable personas, whereas students in the fifth category created personas by relying on their personal experiences. This self-centric strategy can be divided into two sub-strategies: (a) stereotyping strategy and (b) realistic strategy. Students in the sixth category created personas by combining elements from different strategies, e.g., 'data-oriented strategy' and 'imaginative strategy'.

Based on their own accounts, the students typically started their persona creation process using data as the starting point, which involved filtering and analysing the analytics data. The students first paid attention to consumers' spending habits, filtered data based on spending, and divided consumers into high or low-spending consumers. Gender, profession, and age were also mentioned as main characteristics. Based on the data, these characteristics provided an overall picture of the company's customers and reflected most of them.

Most students (n=25, 86.21%) focused on existing data ('data-oriented strategy') at least at the starting point of creating personas. In several cases (n=18, 62.07%), students also used another strategy besides the data-oriented strategy ('mixed strategy'). A few students (n=7, 24.14%) focused only on existing data ('data-oriented strategy') and did not report using other techniques.

Reasoning by Creators

Age: Most of the students (n=23, 79.31%) decided persona ages by filtering the data: they used the most repeated age or the most common age groups. The average age was also mentioned several times. The age data was filtered and calculated using the characteristics the students wanted to represent e.g., high or low spending and/or gender (e.g., P01, P09).

The students wanted persona ages to be consistent and represent the entire customer base or potential customer segments. A few students (n=10, 34.48%) wanted to have more diverse age between personas ('age diversification strategy'). These students wanted different age groups, different population groups, lifestyles, or generations to be represented, e.g., students and pensioners. To enable this 'diversity strategy', they chose personas from different age categories, e.g., two younger and two older personas (e.g., P12).

Gender: In terms of gender, the largest group of students (n=14, 48.28%) decided on persona gender based only on the data, thereby adhering to the 'data-oriented strategy'. This strategy aims to reach and represent as many people as possible (e.g., P18).

The rest of the students (n=12, 41.38%) decided to create personas for both genders; for example, two males and two females, even if they had gone through the data first. The choice was justified based on parity (i.e., 'fairness strategy'), and the fact that these two, male and female, were the available genders in the dataset. Based on the data, some students mentioned that males were the dominant gender, but they still chose both genders for the persona set, wanting to create examples of both genders for both spending categories ('gender balance strategy') (e.g., P24).

Nationality: No geographic restrictions or specific nationalities were given in the data, and it was also not stated in advance whether the company operates in international or domestic markets. Therefore, there is much variation in whether the students focused on just one country, Europe, or the global market (e.g., P22).

Business opportunities justified the choice of certain nationalities, as the students thought the persona would fit a market, and where the service would have been sold the most or the markets would be suitable for the new startup. Global countries were chosen to emphasize how global the application could be ('secondary data strategy'). The students also used nationalities they knew more about, e.g., their own home country ('self-centric strategy'). The students also used their imagination to select the nationality randomly ('random selection strategy'). Three students mentioned that they searched the Internet for additional information (e.g., P03).

Picture: More than half of the students (n=18, 62.07%) selected persona pictures by searching for pictures related to certain characteristics, e.g., age, gender, profession, or nationality, to represent persona personalities and lifestyles (e.g., P18). The rest (n=7, 24.14%) chose the pictures using their imagination: the pictures that corresponded to the description and what the students had in mind or liked most (e.g., P26).

Smiling and friendliness were mentioned as a criterion for a good persona picture ('sociability strategy'), but at the same time, pictures where the person was presented respectfully and pictures that can be taken seriously (e.g., P03). Some students (e.g., P10) mentioned that it was important for them to create diverse personas ('ethnic and cultural diversity').

Name: The largest group of students (n=12, 41.38%) chose the name by country: they searched for the most common names by country or chose names they knew were typical of the country. Age and year of birth also influenced the choice of name (e.g., P18).

A few students reported using pure imagination or creativity in creating the names for their personas (n=7, 24.14%) or choosing names completely randomly (n=3, 10.34%). One student (3.45%) generated names with multiple random generators and chose the names that inspired them most. Some students chose the names of their friends, acquaintances, or

celebrities (n=5, 17.24%). Two students (6.90%) said they had asked for name recommendations from their friends in the persona's country.

Marital Status: Most students (n=24, 82.76%) used the data (i.e., 'data-oriented strategy'). They looked at how many customers were married or not by various characteristics and used the most common or average marital status (e.g., P25), and one student inferred the marital status from the family size ('data associations strategy') (P29).

A few students (n=3, 10.34%) used only "common sense" ('common sense strategy') and made the decision themselves. The students wanted the marital status of the persona to be as realistic and representative as possible ('realistic strategy / real-world strategy'). A few students said they chose several statuses: e.g., married, single, and divorced ('diversity strategy') (e.g., P12).

A couple of students (n=2, 6.90%, e.g., P09) concluded that older personas were more likely to be married than younger personas, this being an example of the common-sense strategy.

Graduation Status: Most of the students (n=24, 82.76%) made an analysis based on the data by looking at averages and other descriptive statistics (e.g., P22).

One student (P13) created more graduation statuses than just graduated or not graduated, while another student (P16) made an interpretation based on persona age and profession ('secondary data strategy'). Similarly, a couple of students (n=2, 6.90%) concluded that the oldest personas were more likely to have a degree and most of the youngest were not, and it was clear that some professions required a degree ('common sense strategy'). P16 also concluded that if the persona has a job, she/he is graduated.

Profession: The most common approach among the students (n=26, 89.66%) was to determine persona professions based on the data and search for the most common professions ('central tendency strategy'). The most common professions were usually filtered by spending or income, and often also by gender ('gender balance strategy') (e.g., P22).

Three students (n=3, 10.34%) decided on the profession alone. The profession was chosen based on what corresponded to the created personas or what seemed more generic than the professions based on the data, though some students (e.g., P12) used an imaginative strategy.

Work Experience: As with the other persona attributes, the interpretations based on the data were the most popular ('data-oriented strategy'), but this time more students used their judgment because they found the data confusing and inconsistent. For example, a 50-year-old lawyer had only one year of working experience, which did not make sense (these artifacts were due to the fictitious nature of the dataset). According to the students (e.g., P09), the average years of work experience did not make sense regarding age, profession, spending, or income.

The most common or average work experience was filtered according to age, profession, and/or spending ('central tendency strategy') (e.g., P20). The students also used their judgment, sense, and logic ('common sense strategy') to determine the persona's work experience, for example, based on the age of the personas (e.g., P28).

Family Size: Most students (n=23, 79.31%) determine the size of the persona's family by selecting the most common value or by computing the average based on the data ('central tendency strategy'). The most common family size was determined by combining certain characteristics, such as gender and spending (e.g., P22). Two students (6.90%) reported

adjusting the average towards diversity ('diversity strategy'). One student assumed that older personas were more likely to have started a family than younger ones ('common sense strategy') (e.g., P01).

The students also wanted to create different family sizes so that it was possible to compare them and see the differences ('comparative strategy') (e.g., P13).

Spending Score: The spending scores were filtered from the data according to profession, age, and/or gender. Four students (13.79%) started the process with this step. They said that spending score was the most critical variable and formed the basis of the personas (e.g., P10). In the assignment, the spending score was determined as an important segmentation factor.

Many of the students (n=11, 37.93%) concentrated only on high and low-spending customers. In several cases (n=6, 20.69%), a symmetrical number of personas were selected from both spending categories, e.g., two high spenders and two low spenders. This decision was justified by the fact that the students wanted to be equal ('diversity strategy'), and it was easier to analyse the differences between symmetrical personas (e.g., P16).

Text Description: When designing the text description of the persona, the students predominantly had to use their creativity ('imagination strategy') because the data was no longer of much help. Although the data provided background information, it did not suggest any new information for the text description (e.g., P11).

The students imagined, for example, what goals or problems the persona working in a certain profession might have ('common sense strategy'). They were inspired by the profile they had just created using the data. They also used stereotypes to create descriptions ('stereotyping strategy'). However, the primary goal was to create a realistic description (e.g., P12). Based on the task description, some students assumed that the personas were active and interested in fitness and/or well-being ('secondary data strategy') (e.g., P15).

The example of a persona text description provided by the educator and the pictures chosen by the students were used as inspiration. A few students (n=3, 10.34%) reported that they came up with hobbies and a suitable description based on the chosen persona picture ('data associations strategy'). The students were also inspired by the people around them, e. g., friends, and family, and by their own experiences ('self-centric strategy') (e.g., P16). One of the students (P03) searched for information and inspiration on the Internet ('secondary data strategy'.

Underlying theoretical aspects for choices in persona creation

We highlight key theoretical aspects of persona creation to strengthen the empirical analysis. First, regarding the representational choices in persona profiles, our findings indicate that student creators often make decisions based on data-driven objective insights and personal subjectivism (including perhaps biases). These decisions (e.g., selection of age, gender, nationality, and profession) are not merely reflective of the data, but are also shaped by the student creators' perceptions of what constitutes a realistic or representative persona. The findings of this research present insights into the collective and performative dimensions of persona creation (Moore et al. 2017; Marshall et al. 2019), particularly evident in how participants draw from their social circles and cultural references when assigning names and visual elements to personas. This implies that there is an objective-subjective dichotomy involved in persona creators' calculus of determining persona attributes, which links to prior literature concerning the less-than-pure objectivity of persona creation processes (B. Jansen et al. 2021; Salminen et al. 2021).

Also, it is evident that the student creators' motivations extend beyond technical accuracy to include values such as fairness, diversity, and social acceptability. The desire to create balanced and diverse personas reflects broader societal values (Wilson et al. 2018; Guan et al. 2023) that creators bring to the persona creation task. This, in turn, appears to shape the personas profiles they produce. This finding aligns with theories of UCD that emphasize the role of empathy and ethical considerations in design processes (Goodman-Deane et al. 2018; Hiléia da Silva Melo et al. 2021).

DISCUSSION

Contribution to persona design theory

The observed richness in persona creation strategies demonstrates the adage that instead of a single 'go to strategy', creators (at least these student creators) opt for multiple analytical paths; these paths are not only based on data but also on value judgments such as gender parity or diversity. In our case, the creators were from the so-called Generation Z cohort (i.e., young people in their early twenties at the time of the study). Their values' impact on the personas could be seen in the reasoning, for example, to create diverse and demographically balanced personas, which may be common values in this generation.

So, even though the persona creation process is often presented in the literature to be based on rigorous, systematic steps that give an appearance of control and standardisation, the reality appears different. Rather than being controlled, objective, and systematic, in the sense of obtaining standard results from the same dataset, there is a great deal of variation and subjectivity in the process, which then manifests into variation in the created personas. In other words, when using manual persona creation, personas are likely to differ even when created from the same data. Persona creators follow idiosyncratic analysis procedures that share some commonality but also a degree of uniqueness. Interestingly, the deviation in the creation process can also reflect the nature of personas: while two personas can be similar, they are seldom exactly alike.

Particularly, two tendencies are striking: a tendency to create personas that are (a) demographically balanced and (b) presented in symmetrical proportions. The persona sets created were much more likely to contain an even number of personas (72.4%) than an odd number of personas (27.6%). Similarly, most participants (65.5%) created the same number of male and female personas. These trends illustrate that persona creators strive to create 'harmonious' user representations from data that necessarily do not contain such harmony on a raw basis. Furthermore, the findings suggest that creators tend to create a relatively small number of personas, possibly limiting a diversified view of the user base. This poses an issue for inclusive design, which is hard to avoid in manual persona creation (Goodman-Deane et al. 2018).

Our main contribution is the taxonomy for persona creation strategies, a starting point for more detailed scrutiny. Insights into this individualistic process have been lacking in the literature, even though it is beneficial to understand how personas are created and why persona creators choose certain information to be included in the persona profiles. This line of study parallels manual and automatic persona generation (An, Kwak, Jung, et al. 2018), especially connected through the concept of 'persona transparency' (Jung et al. 2018). Algorithms can be queried, to some extent, using computational means (explainable algorithms (Garfinkel et al. 2017)), but what is the parallel for manual persona creation? We asked the participants to illustrate their thought processes. This is one evident means to inquire about the process; other means could be further developed.

Practical implications

For persona creators, our findings offer means for self-reflection and identification concerning one's style of persona creation: Which type of persona creator are you? For organizations employing personas in UCD, our findings imply a need to document the persona creation process in greater detail so that stakeholders precisely understand where each attribute originates from. Often, such details are left out, omitting nuances in the creation process while risking that this important 'tacit knowledge' remains uncommunicated to others in the organization.

Organizations deploying personas could opt for an ensemble approach, involving several individuals or teams creating the personas independently from the same data and then reconciling the outputs to mutually agree upon the final set of personas to be presented to the stakeholders. In the process of reconciliating different views, explaining the choices behind the persona information selection (e.g., "Why did you choose this age to Persona 1? Please elaborate your analytical thinking process.") can facilitate others' understanding of the differences and thus help in merging different viewpoints into a set of personas that represent the varied insights that different creators have observed, as well as yielding a persona set that the creators familiar with the dataset can accept. Such a consensus can lead to more robust personas that can be better defended in the organizational adoption and decision-making processes.

Limitations and future work

How individuals analyse user data and form mental models of said data are cognitive processes based on opaque cognitive processes that we cannot directly observe. We interpreted these processes based on what the participants told us, i.e., their thoughts. Future studies could apply methods such as psychological measurement (Sourina & Liu 2014), eye tracking, and think-aloud (Alhadreti & Mayhew 2017) to investigate what aspects of the user data individuals pay attention to when creating personas while simultaneously asking the participants to justify their design choices and, through this manner, learn from how they explain their analytical process. Replication studies with other designers and larger samples would also be fruitful research.

CONCLUSION

This study examines persona creation strategies employed by student creators, highlighting the dichotomy between objective and subjective influences. Six distinct persona creation strategies emerged, contributing to a better understanding of how creators interpret and transform data into personas of user representations. Our findings call for a reflective approach to documenting the rationale behind design choices. The results also advance persona creation theory by bridging the gap between methodological approaches and the personal values that appear to shape persona design outcomes. As, such, this research offers insights for both researchers and practitioners employing personas.

WORKS CITED

Alhadreti, O. & Mayhew, P., 2017, 'To Intervene or Not to Intervene: An Investigation of Three Think-aloud Protocols in Usability Testing', *J. Usability Studies*, 12(3), 111–132.

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), 54.

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(3).

Bødker, S., Christiansen, E., Nyvang, T. & Zander, P.-O., 2012, 'Personas, people and participation: challenges from the trenches of local government', *Proceedings of the 12th Participatory Design Conference on Research Papers: Volume 1* - PDC '12, 91, ACM Press, Roskilde, Denmark.

Brickey, J., Walczak, S. & Burgess, T., 2012, 'Comparing Semi-Automated Clustering Methods for Persona Development', *IEEE Transactions on Software Engineering*, 38(3), 537–546.

Chapman, C. & Feit, E.M., 2019, *R For Marketing Research and Analytics*, Springer.

Chapman, C. & Milham, R.P., 2006, 'The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(5), 634–636.

Cooper, A., 1999, *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*, 1 edition, Sams - Pearson Education, Indianapolis, IN.

Curry, A.C., Robertson, J. & Rieser, V., 2020,;Conversational assistants and gender stereotypes: Public perceptions and desiderata for voice personas; *Proceedings of the second workshop on gender bias in natural language processing*, 72–78.

Damiano, R., Gena, C., Maieli, A., Mattutino, C., Mazzei, A., Miraglio, E. & Ricciardiello, G., 2022, 'UX Personas for defining robot's character and personality', arXiv preprint arXiv:2203.04431.

Emmanuel, G.-S. & Polito, F., 2022, How Related Are Designers to the Personas They Create?, *Design, User Experience, and Usability: Design Thinking and Practice in Contemporary and Emerging Technologies: 11th International Conference*, DUXU 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part III, 3–13, Springer.

Garfinkel, S., Matthews, J., Shapiro, S.S. & Smith, J.M., 2017, 'Toward Algorithmic Transparency and Accountability', *Commun.* ACM, 60(9), 5–5.

Giles, D.C., 2020, 'A typology of persona as suggested by Jungian theory and the evolving persona studies literature', *Persona Studies*, 6(1), 15–29.

Glaser, B.G. & Strauss, A.L., 1967, *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., 2018, Evaluating Inclusivity using Quantitative Personas, In *the Proceedings of Design Research Society Conference* 2018, Limerick, Ireland.

Guan, K.W., Salminen, J., Jung, S.-G. & Jansen, B.J., 2023, 'Leveraging Personas for Social Impact: A Review of Their Applications to Social Good in Design', *International Journal of Human–Computer Interaction*, 0(0), 1–16.

Hiang, G., Kulathuramaiyer, N. & Zaman, T., 2017, 'Riding Waves of Change: A Review of Personas Research Landscape Based on the Three Waves of HCI', in J. Choudrie, M.S. Islam, F. Wahid, J.M. Bass & J.E. Priyatma (eds.), *Information and Communication Technologies for Development*, IFIP Advances in Information and Communication Technology., 605–616, Springer International Publishing, Cham.

Hiléia da Silva Melo, Á., Oran, A.C., Silva dos Santos, J., Rivero, L. & da Silva Barreto, R., 2021, *ACF: An Autistic Personas' Characteristics Source to Develop Empathy in Software Development Teams, HCI International 2021-Late Breaking Papers: Cognition, Inclusion, Learning, and Culture: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* 23, 223–236, Springer.

Jansen, B., Salminen, J., Jung, S. & Guan, K., 2021, *Data-Driven Personas*, 1st edn., vol. 14, Morgan & Claypool Publishers.

Jansen, B.J., Jung, S., Salminen, J., Guan, K. & Nielsen, L., 2021, 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'21), 10, Virtual conference.

Jansen, B.J., Jung, S.-G., Nielsen, L., Guan, K.W. & Salminen, J., 2022, 'How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment', *Pacific Asia Journal of the Association for Information Systems*, 14(3).

Jung, S.-G., Salminen, J., An, J., Kwak, H. & Jansen, B.J., 2018, Automatically Conceptualizing Social Media Analytics Data via Personas, *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018)*, San Francisco, California, USA.

Karahasanović, A., Følstad, A. & Schittekat, P., 2021, Putting a Face on Algorithms: Personas for Modeling Artificial Intelligence, *Artificial Intelligence in HCI: Second International Conference, AI-HCI 2021*, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, 229–240, Springer.

Long, F., 2009, Real or imaginary: The effectiveness of using personas in product design, *Proceedings of the Irish Ergonomics Society Annual Conference*, vol. 14, Irish Ergonomics Society Dublin.

MacDonald, C.M., Rose, E.J. & Putnam, C., 2022, 'How, Why, and with Whom Do User Experience (UX) Practitioners Communicate? Implications for HCI Education', *International Journal of Human–Computer Interaction*, 38(15), 1422–1439.

Marshall, P.D., Moore, C. & Barbour, K., 2019, *Persona studies: an introduction*, John Wiley & Sons.

Märtin, C., Bissinger, B.C. & Asta, P., 2021, 'Optimizing the digital customer journey—Improving user experience by exploiting emotions, personas and situations for individualized user interface adaptations', *Journal of Consumer Behaviour*.

Melo, Á.H. da S., Rivero, L., Santos, J.S. dos & Barreto, R. da S., 2020, PersonAut: a personas model for people with autism spectrum disorder, *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, 1–6.

Moore, C., Barbour, K. & Lee, K., 2017, Five dimensions of online persona, *Persona Studies*, 3(1), 1–11.

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

Nielsen, L., 2019, *Personas - User Focused Design*, 2nd ed. 2019 edition, Springer, New York, NY.

Nielsen, L., Hansen, K.S., Stage, J. & Billestrup, J., 2015, 'A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations', *Int. J. Sociotechnology Knowl. Dev.*, 7(1), 45–61.

Nielsen, L. & Storgaard Hansen, K., 2014, Personas is applicable: a study on the use of personas in Denmark, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1665–1674, ACM.

Nolte, A., Lueneburg, K., Wallach, D.P. & Jochems, N., 2022, Creating Personas for Signing User Populations: An Ability-Based Approach to User Modelling in HCI, *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, 1–6.

Pruitt, J. & Adlin, T., 2006, *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*, 1 edition, Morgan Kaufmann, Boston.

Pruitt, J. & Adlin, T., 2010, The Persona Lifecycle, 1st edition, Morgan Kaufmann.

Rapp, A., 2019, 'Design fictions for behaviour change: exploring the long-term impacts of technology through the creation of fictional future prototypes', *Behaviour & Information Technology*, 38(3), 244–272.

Salminen, J., Guan, K., Jung, S., Chowdhury, S.A. & Jansen, B.J., 2020, A Literature Review of Quantitative Persona Creation, *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14, ACM, Honolulu, Hawaii, USA.

Salminen, J., Guan, K., Jung, S.-G. & Jansen, B.J., 2021, 'A Survey of 15 Years of Data-Driven Persona Development', *International Journal of Human–Computer Interaction*, 1–24.

Salminen, J., Rao, R.G., Jung, S., Chowdhury, S.A. & Jansen, B.J., 2020, Enriching Social Media Personas with Personality Traits: A Deep Learning Approach Using the Big Five Classes, in H. Degen & L. Reinerman-Jones (eds.), *Artificial Intelligence in HCI, Lecture Notes in Computer Science.*, 101–120, Springer International Publishing, Cham.

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*.

Seidelin, C., Jonsson, A., Høgild, M., Rømer, J. & Diekmann, P., 2014, 'Implementing Personas for International Markets: A Question of UX Maturity', *Proceedings at SIDER'14Royal Institute of Technology*, KTH, Stockholm, Sweden.

Siricharoen, W.V., 2021, Using empathy mapping in design thinking process for personas discovering, Context-Aware Systems and Applications, and Nature of Computation and Communication: *9th EAI International Conference, ICCASA 2020, and 6th EAI International Conference, ICTCC 2020*, Thai Nguyen, Vietnam, November 26–27, 2020, Proceedings 9, 182–191, Springer.

Sourina, O. & Liu, Y., 2014, EEG-enabled Affective Human-Computer Interfaces, in C. Stephanidis & M. Antona (eds.), Universal Access in Human-Computer Interaction. *Design and Development Methods for Universal Access, Lecture Notes in Computer Science.*, 536–547, Springer International Publishing, Cham.

Tarkkanen, K., Ravyse, W., Katajapuu, N., Saarenpää, T. & Luimula, M., 2018, Applying personas in designing exergames for knee arthrosis, *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–5.

Turner, P. & Turner, S., 2011, 'Is stereotyping inevitable when designing with personas?', *Design studies*, 32(1), 30–44.

Villareale, J., Harteveld, C. & Zhu, J., 2022, '"I Want To See How Smart This AI Really Is": Player Mental Model Development of an Adversarial AI Player', *Proceedings of the ACM on Human-Computer Interaction*, 6(CHI PLAY), 1–26.

White, P.J. & Devitt, F., 2021, 'Creating Personas from Design Ethnography and Grounded Theory.', *Journal of Usability Studies*, 16(3).

Wilson, A., De Paoli, S., Forbes, P. & Sachy, M., 2018, 'Creating personas for political and social consciousness in HCI design', *Persona Studies*, 4(2), 25–46.