

Optimizing the Use of the Sentence Completion Survey Technique in User Research: A Case Study on the Experience of E-Reading

Carine Lallemand

University of Luxembourg & Eindhoven University of Technology

Emeline Mercier

École Normale Supérieure de Lyon

ABSTRACT

Sentence completion, originally a semi-projective psychological technique, has been used as an effective and lightweight user research method in UX design. More information is yet still needed to understand how different sentence stems probe users' insights, thereby providing recommendations for effective sentence completion surveys. We used the completion method on a large-scale sample to explore (e-)readers' experiences and needs. Depending on their reading habits, participants (N=1880) were asked to complete a set of sentences, as part of a web survey. With 14143 user ideas collected in two weeks, our results confirm that remote online sentence completion is a cost-effective data collection method able to uncover feelings, attitudes, motivations, needs, or frustrations. Variation in sentence stems affected collected data in terms of item response rate, idea quantity as well as variety and originality. Building on previous research, this paper delivers actionable insights to optimize the richness of sentence completion outputs.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); User studies.

KEYWORDS

sentence completion, user research method, user experience, online survey, e-reading

ACM Reference Format:

Carine Lallemand and Emeline Mercier. 2022. Optimizing the Use of the Sentence Completion Survey Technique in User Research: A Case Study on the Experience of E-Reading. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*, April 29–May 05, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3491102.3517718>

1 INTRODUCTION

Please complete the following sentence: “After having read the abstract of this paper, I expect ____” This incomplete sentence is an example of a sentence stem used in the sentence completion technique. Originally used in psychology studies [12, 18, 23, 44, 46], the sentence completion technique (SCT) has recently been transferred

to the field of user experience (UX) design where it is used as a user research method to elicit user feedback [8, 30, 31, 68].

To account for the richness and complexity of experiences, researchers have responded to the challenges underlying UX by developing new methods or by adapting existing ones. Those methods are very diverse as they draw on a multiplicity of perspectives and backgrounds associated with UX research, i.e., the fields of design, psychology, or social sciences to name the most prominent ones. A majority of UX methods are still in the early phases of development and their validity has not yet been fully examined [32, 67]. They thus need to be further developed to maximize their potential and scientific quality.

Sentence completion combines a projective technique and questionnaire whereby respondents are provided with the beginnings of sentences that they complete in ways that are meaningful to them [57]. Previous studies in UX have shown the relevance of SCT to explore users' needs or to evaluate UX [30, 31, 68]. One of the main interests of the technique also lies in the ability to collect rich qualitative data remotely and in a cost-effective way [68]. However, while it appears as a valuable and promising asset to extend the methodological toolkit of both scientists and practitioners, very few publications in HCI inform on the use of SCT. Much of the research using this technique up to now has been conducted by a single research team, and one counts less than 20 publications related to SCT in total in our field. Very little knowledge thus exists on best practices to optimize the use of SCT for user research.

In the present paper, we used the SCT on a large-scale sample (N=1880) to explore (e-)readers' experiences and needs. By analyzing how the variation in sentence stems affects the quantity and quality of collected data, we provide key insights and actionable recommendations to optimize the use of the sentence completion method. We also discuss the domain-specific insights gained from the study with regards to previous literature on e-books and e-reading to reflect on the validity of our findings and their practical relevance to better understand e-reading experiences. Our research can benefit both researchers and practitioners, who often need lightweight methods to collect rich user feedback. To the best of our knowledge, there is no comparable study investigating the sentence completion method for user research in such depth (neither in terms of sample size, numbers of ideas collected, or type of data analysis conducted). With this paper, we also contribute to the consolidation of user research methods, which are often simply considered a means to an end (of designing and evaluating experiences), and call for more methodological work in our field.



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

CHI '22, April 29–May 05, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9157-3/22/04.

<https://doi.org/10.1145/3491102.3517718>

2 RELATED WORK

2.1 User Research Methods: Current State, Development Needs and Relevance to Practice

In the UX design process, user research methods are used to get insights into users' needs. By providing an overview of existing UX evaluation methods, Vermeeren et al.'s paper [67] allowed taking stock of the state of UX research and highlighting what additional types of methods and tools were needed. Methods to study the anticipated user experience, the temporal dynamics of UX, and lightweight methods in general were defined as the priority focus for the community. According to Stolterman [59], research frequently fails at supporting practice because "it has *not* been grounded in and guided by a sufficient understanding and acceptance of the *nature of design practice*." So, despite the disposition of UX to be an applied field of research (i.e., to generate information that can be directly applied to real-world problems; [3]), the methods brought by academia do not often satisfactorily address the needs of industry. In recent years, academics attempted to develop methods that are flexible, applicable, cost-effective, and lightweight in order to support product development [27, 28]. However, by doing so, they sometimes have to reach a challenging compromise between the scientific quality of a method and its practicability [67]. Regarding the sentence completion method, Kujala et al. [31] stated that their "goal was to develop a practical technique for gathering qualitative user feedback for product development purposes rather than to develop a strict measurement tool. Sentence completion was selected as it appears practical to use in product development contexts." [31] (p. 12) The authors recognized that systematic research is needed to ensure the validity and reliability of the sentence completion approach. Yet, with the goal "to develop a practical evaluation method for design purposes, we can question to what extent these criteria are applicable" ([31] p. 14).

2.2 The Sentence Completion Technique (SCT)

The SCT combines a projective technique and questionnaire whereby respondents are provided with the beginnings of sentences that they complete in ways that are meaningful to them ([57] p. 132). It is based on the common principle of "fill in the blanks". The SCT originates from the field of psychology, initially introduced by Ebbinghaus as part of an intelligence test [12]. Nowadays, sentence completion tests are among the most popular personality assessment instruments used by practitioners [18, 23] for clinical applications but also in management [6] or education [2]. As the technique is relatively new in UX research, knowledge from psychology is relevant to consider. We will describe insights from experimental studies in section 2.3.

Since the method has been imported from the field of psychology to the field of UX around 10 years ago by Walsh, Nurkka, and Walsh [70], around 20 published studies only have used the sentence completion method to collect users' insights, the large majority of them being published by a single research team. While this is a great indicator of the research continuity and quality at this specific institution, it also echoes Harrison and Tatar's metaphor [20] on methods being like toothbrushes. "Everyone uses them, but no one

likes to use someone else's". Long-term methodological consolidation efforts should ideally be distributed over the community since the scientific process implies building on previous research to produce cumulative knowledge [32]. In the case of the SCT, there is still a dearth of methodological research to compare the insights generated by the method to other methods in user research or to optimize the use of the methods in user research studies.

The method has majorly been used to evaluate user experiences and to explore users' needs and values. Studies cover different types of interactions with products or services, such as activity trackers [30, 69], smartphones [31, 70], or games [29]. The administration of the sentences to complete is done face-to-face during an interview or remotely under the form of an online survey. The sample sizes in these studies vary from N=10 [29] to N=104 [69]. On a more anecdotal note, sentence completion has also been used as material to familiarize workshop participants with the theme of the research [66], as part of a cultural probe study [64] or diary study [40]. Interestingly, Rohles et al. [48] adapted the technique in the form of a fill-in-the-blank concept map.

Numerous benefits have been highlighted in previous studies: the sentence completion method is less time-consuming than an interview, yet more informing than a multiple choice questions survey. As respondents use their own words to describe their situation, "they might thus give more spontaneous and honest answers as compared to traditional questionnaires" [25]. This absence of constraints could be a major asset of the sentence completion method, overcoming the limitations of UX standardized questionnaires, whose items are sometimes described as "awkward", "not suitable for the context" or "not able to fully account for their experiences" [34, 35]. The sentence completion method could thus be an alternative way of assessing UX without constraining the user by a predefined vocabulary like the one typically used in standardized scales. A mode widespread use does however call for more knowledge about best practices to construct, implement and analyze sentence completion surveys. Unfortunately, very little knowledge exists on best practices to optimize the use of the sentence completion survey method for user research. Selection of items (if one wants to design a standardized test), stems length, stems affective tone, administration instructions, or scoring are seldomly described. As an example, a recent study uses the sentence completion technique to build a standardized questionnaire assessing sleep experience [17]. Unfortunately, the construction of the questionnaire is not well detailed and it is thus hard to assess whether the selection of items, the length of sentence stems, their affective tone or the position of the blank to be filled have been considered in the design of this tool.

2.3 Variations in the Approach to the Sentence Completion Technique

Existing studies have very briefly listed some good practices about the application of SCT, summarized in a chapter of the UX methods textbook by [33]. The study of Walsh, Nurkka, and Kujala [69] highlights, for instance, the effect of language variations in sentence stems, by reporting findings on an SCT survey completed by either native or non-native speakers. Unsurprisingly, native speakers were better able to express their experiences, needs and emotional states and gave longer and more detailed answers. Besides that, papers

using SCT somehow lacked transparency on revealing the sentence stems used, or depth about the way data is reported.

As the few studies using the SCT in the field of UX do not focus on methodological aspects related to the type of sentence stems used, it is worth looking back at the use of the SCT in the field of psychology. Experimental studies with varying conditions of sentence stems were conducted as early as 1940s [51, 58, 62, 65]. Authors tested the impact of different variables on the quality and validity of participants' responses, with the main focus on the use of a first-person versus third-person perspective [42, 51], the tone [39], the length of the sentence stem [65], or the accompanying instructions [9]. Murstein et al. [42] for instance tested variations in stem subject (first-person vs. impersonal), as well as affective tone (positive, negative, neutral). First-person stems elicited more projection than impersonal stems. Furthermore, neutral stems showed a higher projection level as compared to positive or negative stems. Regarding the length of the stimulus, Weisgerber [71] assumed that the longer the stem, the more the participant is steered to a certain dimension. Longer stems would lead respondents to more specific types of responses. This, in turn, would also ease the interpretation of the item by the experimenter, as unstructured items can produce a wider variety of responses that are harder to interpret [44]. Additionally, Rogers, Bishop, & Lane [46] claim that the length of the stem, together with the length of the total test, affects the motivation of the participant, assuming that more stems to complete will yield to fatigue.

3 RESEARCH OBJECTIVES

Our objective is to consolidate and deepen previous work on the sentence completion method, by analyzing sentence stems outputs on a large sample through a real-life case study. We did not intend to compare sentence completion outputs to the ones obtained through any other user research technique but rather wanted to compare several variants of sentence completion items. Understanding how to best formulate sentence stems in sentence completion surveys will eventually support the optimization of the use of this method in the UX design field. Our research questions are the following: How does variation in sentence stems affect collected user data? Which sentence stem formats, amongst the ones included in the study, are the most effective to gather rich user insights using the sentence completion method?

We examined the impact of four variations in sentence stems using three metrics described in the methodological section. We derive actionable guidelines for researchers and practitioners willing to use the online sentence completion method and to optimize the richness of collected data.

4 METHOD

We collected data on sentence completion outputs through an online case study on the experience of digital books. Despite their potential to shape novel reading experiences, several factors seem to inhibit the wider adoption of electronic reading devices. The project's stakeholders were therefore interested in understanding users' experiences, needs, expectations, or frustrations. However, while the rich feedback gathered from users served for design purposes to improve e-reading experiences, the project was from our

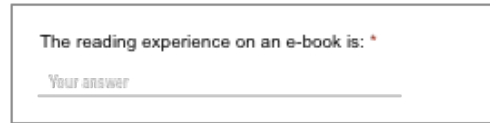


Figure 1: Display of the sentence stems on Google Forms

perspective an opportunity to conduct the present methodological study. We were therefore not interested in the content of user ideas per se, but in their nature in terms of quantity, variety and novelty, and in how the method compared to previous work in terms of domain-specific insights collected. We included different sentence stems in the survey in order to understand the nature of answers prompted by each type and to eventually derive recommendations for the design of sentence completion surveys. Using a real-life project allowed us to increase the ecological validity of our findings. However, we had to make a few trade-offs related to the number of sentence stems to experiment in order to keep the online survey realistically short (completion time below 15 min). Overall the sentences covered all aspects of interest to the project's stakeholders and presented sufficient variations to address our research questions.

4.1 Participants

1880 people participated in the survey (69.6% females, 30.4% males). Only fully completed responses were included in the analysis. Empty or incomplete responses were not saved by the system. No compensation was provided for participation. The mean age of the sample was 35.7 years (Min = 11, Max = 80, SD = 13.3). Amongst respondents, 1284 (68.3%) declared reading e-books and were assigned to the readers' path (coded as R) through conditional branching. The remaining 596 (31.7%) declared reading print books only and were therefore assigned to the non-readers path (coded as NR) through conditional branching in the survey.

In our sample, 6% of participants declared having a disability likely to impact their reading experience, mainly vision deficits. Using a convenience sampling strategy (skewed with regards to the language spoken by participants), our sample nevertheless covers employment categories: 30.7% of respondents occupy a managerial position, 19.2% are employees or workers, 22.6% were students, 5% were retired and 8% were unemployed. Regarding reading habits, 57% of respondents reported reading at least a book per month.

4.2 Procedure

The survey was broadcast online to multiple advertisement channels for two weeks. To access a diverse population, social networks, mailing lists, and forums on the topic of reading or online reading clubs were targeted as the main vector of dissemination. The sentences to complete were included in an online survey using Google Forms (under the form of short free text fields) (Figure 1).

All sentences were displayed on a single screen and in the same order. These were defined as mandatory questions. The sentences were adapted to the respondents' profile using conditional branching, thus creating a custom path through the survey based on the answer to a screening question. Non e-book readers (declaring that

Table 1: Sentence stems and associated dimensions prompted. The order of sentences follows the one of the survey.

Users' profile	Code	Sentence stem	Dimension
Non e-book readers (n=596)	QNR1	Compared with a print book, an e-book is. . .	Comparison between products
	QNR2	In my opinion, e-books are targeted to. . .	Identity / product image
	QNR3	I have never read any e-books because. . .	Frustrations / Barriers to use
	QNR4	I would read an e-book if. . .	Expectations and needs
	QNR5	I expect an e-book to. . .	Expectations and needs
	QNR6	When I read a print book, I feel. . .	Affects
E-book readers (n=1284)	QR1	Compared with a print book, an e-book is. . .	Comparison between products
	QR2	The reading experience on an e-book is. . .	Global UX
	QR3	The problem with e-books is. . .	Issues and frustrations
	QR4	What I love about e-books is. . .	Positive aspects
	QR5	What frustrates me the most with e-books is. . .	Issues and frustrations
	QR6	I find that the interface of an e-book is. . .	Specific UX – interface
	QR7	I dream of an e-book that. . .	Expectations / dreams

they only read print books or barely read at all) were invited to complete 6 sentences, while e-book readers (declaring that they at least read e-books sometimes) were presented with 7 sentences. For the sake of brevity, we will refer to non e-books readers as non-readers (NR) and e-books readers as readers (R). The following instructions were displayed before the sentence stems: “Please, complete the following sentences. Respond rather quickly without thinking too long”. The instructions included no indication about the number of words or ideas that respondents should provide.

4.3 Material

4.3.1 Survey Design. We created a total of thirteen sentence stems related to the experience of (e)reading. Sentences were pilot tested ($N = 15$, convenience sampling) prior to the administration of the survey to ensure that each sentence stem would be understandable and would prompt the intended dimension. Cognitive interviewing was used to understand how our respondents processed thoughts and perceptions when completing each sentence stem. The pilot test highlighted the importance of creating an initial pool of several stems per intended dimension as slight changes in the formulation greatly impacted the meaning and the richness of the replies provided by participants. Similar to the creation of a summated rating scale, the pool of items is then reduced to the best sentence stems only. Jargon words were removed and the most understandable stems only were kept for the study. The wording “interface” was not detected as problematic at this stage, due to biases in the convenience sample used in the pilot test. Results interestingly show how this impacted stem QR6.

All questions were originally in French. Table 1 presents the 13 sentence stems used in our study, along with the underlying dimension which we intended to prompt using each item. Replies were always expected in the target position (i.e. at the end of each sentence stem).

The survey also included demographic questions (gender, age, educational level, socio-economical class, disabilities) and questions about reading habits. Overall, the survey entailed 24 questions in total for e-books readers and 19 for people not reading e-books. Because we used sentence completion as part of a wider survey in

a real-world project, we were limited in the number of sentence stems that could potentially be tested. This however increased the ecological validity of our dataset.

4.3.2 Variations in Sentence Stems. Previous literature in psychology [9, 39, 42, 46, 51, 65] focused on experimental variations in the stems subject, the affective tone, the length of sentence stems, the position of the blank, the length of the test, or the accompanying instructions. As our goal is not to replicate these studies but to contribute to the optimization of the method in the context of user research, we decided to instead build on current qualitative studies in UX design to select the variations in sentence stems to be included in our study. To do so, we collected typical questions asked during user research interviews or open-ended questionnaires aimed at understanding users' experiences with a product. In most cases, the interview goes from generic to specific, and follow-up questions often ask the users to clarify or complement their answers. Comparison, analogies, or similar “out of the box” activities are used to stimulate the users. We also considered previous sentence completion surveys used in UX research. We translated these practices in the following variations in sentence stems, which were included in the survey. We analyze and discuss their effect on the data collected in subsequent sections of the paper.

- *Generic vs specific dimensions:* the design of sentence completion surveys leaves a lot of freedom to the researcher. Previous studies usually mixed sentence stems that were very broad (exploring the overall user experience) with stems prompting users to comment on specific dimensions such as aesthetics or credibility for instance. We included QR2 (“The reading experience on an e-book is. . .”) as a global UX measure and QR6 as a specific item focused on the user interface (“I find that the interface of an e-book is. . .”).
- *Comparative sentence stems:* these sentences were formulated as comparisons (in this case between print books and e-books). Previous research has shown the influence of past experiences on UX [49] and the power of analogical thinking. We thus included two comparative items (QNR1 and QR1) as onboarding questions common to the two user profiles.

- *Redundant sentence stems*: in this variation, we probed the same dimension twice through two distinct sentences. As improving the experience was the main concern of the project's stakeholders, we decided to focus on eliciting users' frustrations. The two redundant stems were "The problem with e-books is ___" (QR3) and "What frustrates me the most with e-books is ___" (QR5). Understanding to which extent new user insights can be collected through the use of redundant stems is essential if one aims at informing a specific UX dimension in-depth.
- *Extreme sentence stems*: as previous studies on creativity highlight [52, 54], the originality of a stimulus might prompt ideas that would otherwise not be elicited through more ordinary items. We therefore included an original projective sentence stems asking users' to describe their ideal digital book, "I dream of an e-book that ___" (QR7). While it might seem a bit overrated, we will use the word "extreme" to qualify this original stem to avoid confusion with the originality metric.

4.3.3 Metrics for Assessing Sentence Completion Outputs. Three major metrics, borrowed from creativity studies [52, 54], were used to assess the richness (from both a quantitative and qualitative perspective) of users' replies to our sentence stems: quantity, variety, and originality.

Quantity is commonly defined as the number of generated insights or ideas. In creativity studies, more ideas increase the chance of producing better concepts. In user research, quantity is important to cover the whole range of users' needs, motivations, frustrations, and expectations. We thus analyzed the item response rate and differences in the number of insights per sentence stem and used them as metrics representing quantity.

Variety is commonly defined as a measure of explored solution space during the idea generation process [52]. During user research, collecting similar user insights can indicate that saturation of data has not been reached and can eventually prevent generating quality ideas later in the design process. We used the number of answer categories as an indicator of the richness of the outputs collected in terms of variety.

Originality: The originality of insights or ideas is often assessed by judges who are experts on the topic of interest [50]. However, this approach is hardly scalable to the 14,000 ideas collected through the present study. We therefore used the originality metric defined by So and Joo [54] following the scoring procedure of Guilford's Unusual Uses Test [19]. We measured originality (also called *novelty* [52]) by comparing each idea to the frequency of its occurrence across participants. An idea mentioned by only 5% of respondents is considered 'unusual', and an idea mentioned by only 1% of respondents is deemed 'unique'.

4.3.4 Statistical Analyses. We used SPSS software to perform statistical analyses. As there were no real missing data, the sample size is always $N=1880$, with subsamples of $n=596$ and $n=1284$ for groups of non-readers (NR) and readers (R) respectively. Qualitative answers were coded by the two authors of this paper using an inductive coding approach. The coding categories emerging from the analyses of a sample of 10% of data for each item, independently coded by the two researchers, were discussed between the authors

until agreement was reached. The resulting coding schemes for each item were used to code the entire dataset. The categories identified and occurrences per category are provided for each item in Appendix A. We illustrate some of these categories in the results section using participants' verbatims, translated from French to English.

5 RESULTS: THE EXPERIENCE OF E-READING

5.1 Non E-Books Readers: Attitudes, Resistance Factors, and Expectations

Comparing e-books to traditional books (QNR1), the majority of respondents from the non-reader group (NR) describe e-books in a negative way (82.4% using a negative term, see Table 6). About a third of NR occurrences (30%) describe e-books as something cold, impersonal, without a soul (other terms frequently used include "less intimate", "less alive"). The absence of pleasure is evoked in 12.5% of occurrences and the absence of sensuality (missing touch, smell, or sound of paper) in 10.9%. Most answers are relatively short but include synonyms (e.g. cold + impersonal) and only a few participants explain their feelings. A minority of replies (considered unusual or unique ideas) include the fact that e-books are "not a book" ("it is not a book", "for me it is another product") or provoke the "death of paper books and treasured bookshops".

Asked whom they think e-books are targeted to (QNR2), NR most frequent responses cover travelers (16%), everybody (16%), young people/students (14%), people with tech knowledge (13%). People who read a lot, who do not have storage space, who are in a rush, or who do not like to read complete the picture. On the unusual side, some respondents mentioned "those who do not like to keep a book after reading it", "lazy people" or "who do not find paper books in their native language", to only mention a few examples. Explaining the reasons why they never read e-books (QNR3), NR main replies deplore the lack of haptic and sensual elements (24%) and give a large number of examples illustrating the feeling they miss in e-books: "I like the sensation of paper", "I like to hold a book in my hands, admire its cover – sometimes embossed, to smell its pages and put a bookmark in it", "I like skimming through a book to get an impression of the content", "I feel the need to create a bond with something material when I immerse myself in a story". Other main categories of replies include an undefined preference for paper books (21%), a lack of comfort to read on a screen (18%) – some people highlighting that they already work on screen all day long, and a lack of willingness or interest to read e-books (12%). Unusual replies include "not being able to brag with our bookshelves", "knowing that I own the book, I'm not sure if my e-book will still be readable in 20 years", or the fact that "it might become a habit after all just like I do not use a paper dictionary anymore". Issues of memorability and immersion are also present.

Our next question QNR4 inquired under which conditions NR respondents would read an e-book. Replies showcase a strong resistance from that group, 24% of occurrences mentioning that they would read an e-book if they had no other choice or would be forced to, 16% that they would never read one or are not interested and 13% of mentions highlight that they would do so if the book does not exist in a printed format. On a more positive note, 14% of replies

state that they would read an e-book if they had the device or the file, if they travel (7%), if they would be motivated to do so (6%) and if the business model would change (6%). Contextual elements, for instance related to the format of the book, are visible in the verbatims: “If it’s a book for work that I don’t consider important enough to have the printed version”, “if it is a cartoon because e-books bring added value there”.

Question QNR5 gathered NR’s expectations towards digital books. If 24% of occurrences mention not expecting anything (and 4% even mentioning that they expect e-books to disappear), the question of price came as a second argument (11%), the attractiveness (10%) and ease of use (9%) next. Less common replies (5%) include an enriched experience. These are particularly insightful for designers, yet very few NR mention some creative ideas for instance “I expect an e-book to transport me in the universe of the book”, “to offer more than the text, like bonus content, documents around the book, the author, just like bonus content in DVDs”. Finally, with QNR6, we expected to collect the experience NR strive for, by prompting them to share how they felt when reading a paper book. A third of occurrences (32%) simply mentioned feeling good or happy, another third (29%) conveyed the idea of feeling immersed or absorbed. Relaxed (17%) and free (5%) came next. Most of these replies are a single word or expression but some bring a sense of poetry to express their love for books.

5.2 E-Books Readers: Satisfaction, Frustrations, and Expectations

Comparing e-books to traditional books (QR1), readers (R) are majorly positive (77%), with the main arguments being pragmatic. 40% of occurrences focus on the ability to transport e-books easily (“I can easily have hundreds of books in my table but only a few in my bag”, “I always have all my books in one hand”), 25% describe it as “pragmatic” without further explanation, price (7%) and features or accessibility reasons (6%) being also mentioned as assets. 12% of occurrences nevertheless mention that e-books are less pleasant or attractive than their paper counterparts, for similar reasons as the ones evoked by the NR group (e.g., pleasure derived from haptic elements). Unusual insights describe e-books as “merely a reproduction of the paper metaphor”. About their experience with e-books (QR2), a majority of replies describe it as agreeable and pleasant (60%) whereas 10% refer to it as unpleasant. A quarter of replies compared experiences of e-books to paper books, describing it as less good (9%), similar (7%), better (5%), or different (4%). If respondents sometimes explain already what makes the experience pleasant for them, they elaborate largely on these satisfaction factors in QR4. Portability and storage facility again comes as the main satisfaction factor (45%). The easiness to download or buy e-books comes second (e.g., “I can buy it in one-click at 3 am”) yet with no single mention of a specific platform. Conversely, replies were detailed about the type of features respondents appreciated (10%), with about 20 different features spontaneously emphasized in the replies (the dominant one being the dictionary, full-text search, annotations, bookmark, and backlight).

Next, items QR3 and QR5 both provided interesting insights into what readers considered a problem (QR3) or source of frustration (QR5). In both questions, missing haptic elements dominate the

number of occurrences (15% and 28% respectively) with a larger part in QR5. The importance of physical presence (as a reminder on a nightstand, as aesthetical elements on a bookshelf) is commonly visible in the answers. Price is overall mentioned in 20% of occurrences, with some facts as illustrations “it is too expensive as compared to a print book. Also, when the book is published in a pocket format, the price decreases substantially but not the price of the e-book, why is that?”. The small price difference between e-books and paper books is considered inadequate. Bestsellers and new books are frequently mentioned as particularly expensive. Technical restrictions, especially Digital Rights Management (DRM, a form of copyright protection) are another consequential source of frustration as they challenge the idea of ownership of the book: “what I buy belongs to me and should not disappear simply when I cross a border because Apple or Google decided it”, as well as social aspects “DRMs that prevent me from lending a book to someone”. Social aspects were more frequently mentioned as a source of frustration in QR5 than a problem in QR3. They offer some interesting insights: “There is a symbolic loss: offering a book as a personal gift”, “The problem is that we cannot enjoy the book cover as artistic objects and also cannot get an autograph”, “The absence of physical traces of use which sometimes tell something about previous readers”.

QR6 inquired about the interface and was not well understood by some respondents. Half of them simply declared that the interface was clear or likable, whereas 17% expressed disappointment. It is the question for which respondents most often specified the type of device, format, or feature they address in their reply. Variability between devices, platforms, editors, and books is commonly mentioned as high. Finally, what do readers dream of when it comes to a digital book? (QR7). With only a 75% response rate, this seemed to be harder than expected to envision an ideal artifact. The insights collected were however richer (both in a practical and futuristic sense) to be used as ideation material. The majority of ideas (23%) involve additional features and enhanced interactivity. This is followed by better technical features (13%), such as screen quality or battery life, and lower price (12%). Many haptic elements are again mentioned here, as a comparison to paper books (9%) or as an enhanced feature, e.g., “use different scents along with the story”, “plays music”, “triggers the five senses”. Most ideas remain quite practical, often preceded by an expression that the reader does not have dreams or ideas, “No idea! That gives me the number of pages I have read rather than the percentage of completion?”. Others mention inter-operability with paper books “that paper book would be integrated as a digital support: I could start with the paper book, switch to the tablet, perhaps to my phone and going back to paper. The verb “dreaming” is important”. Adjusting the content according to people’s mindset is another dream: “the device would read my mind and suggest books accordingly”, “the content would evolve according to my wishes and moods”. Or discovering more about the author’s universe: “We could see the text and corrections as written by the author to understand their process”, “We could contact the author directly”, “I could participate not as a reader but as a co-author”.

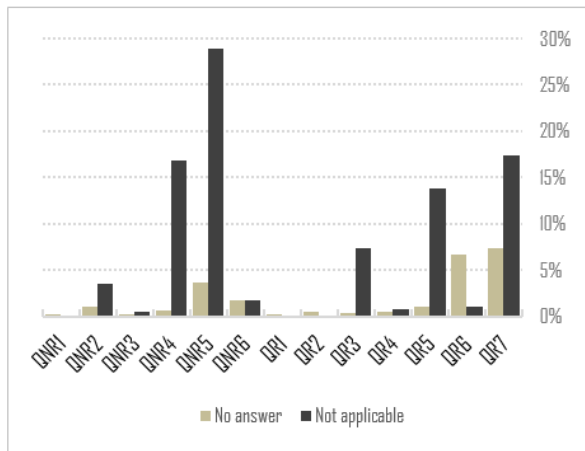


Figure 2: Distribution of empty responses at the item level

6 FORMAL ANALYSIS OF THE SENTENCE STEMS

6.1 Quantitative Metrics

6.1.1 Item Response Rate. Overall, our set of sentence completion items achieved a very high item response rate of 91% on average, without significant differences between readers and non-readers. Out of the 9% of missing data, 1.8% were actual non-responses, and 7.2% were ‘not applicable’ cases. In Figure 2 and Table 2, we distinguish between these missing answers (being unable to leave an answer blank, users usually typed a random character) and “not applicable” (N.A.) answers (where users explained why the question does not match their situation). Typical examples of N.A. answers in our survey occur when a sentence stem assumes users to have expectations about e-books when they actually do not expect anything.

At the item level, the response rate varies between 99.9% (QR1) and 67.4% (QNR5). Comparative sentence stems collected the highest item response rate with 99.7% for QNR1 and 99.9% for QR1 respectively. The lowest item response rates correspond to items QNR5 (67.4%), QR7 (75.3%), and QNR4 (82.5%), which are all prompting respondents about their expectations. This also applies to our extreme item QR7 inviting users to envision a dream e-reading experience. Regarding redundant sentence stems, results show an increase of missing or not applicable answers from QR3 (7.8%) to QR5 (14.9%). Generic sentence stems such as QR2 (99.5%) or QNR6 were also answered by the vast majority of our respondents (96.6%).

6.1.2 Number of Users Ideas Collected. In two weeks, we collected a total of 14143 user ideas generated by 1880 participants through the completion of sentences. In three-quarters of the cases (75.3%), respondents provided a unique idea to each sentence stem. But sometimes they completed the sentences by giving several answers (e.g., “what I love about e-books is that I can carry them easily with me when traveling. And also that they are a bit cheaper than print books”). They gave 2 ideas in 11.7% of the cases and 3 or more ideas in 3.8% of the cases only (Figure 3). The average number of ideas generated per sentence stem is therefore 1.12 (SD = 0.66) with a

maximum of 11 ideas produced on a single question (QR4). As a reminder, the instructions included no indication about the number of words or ideas that respondents should provide.

Interestingly, readers were more prolific when prompted to describe what they love about e-books (QR4) whereas non-readers tend to give more answers to explain the reasons why they do not use e-books (2 or more ideas in 24.3% of cases on QNR3). We conducted one-sample t-tests in order to see how each sentence stem compares in terms of the number of ideas to the mean of our sample (N=1880, M=1.12). Comparative sentence stems scored significantly higher than the mean of the sample on the number of ideas collected, especially amongst readers. On QR1 (QR1, M=1.32, SD=0.64), results show a mean difference of 0.20, 95% CI [0.17 to 0.24], $t(1283) = 11.36, p < .000$.

The *Extreme item* QR7 produced the smallest number of ideas on average (M=0.86, SD=0.62) for readers. It was less likely to trigger more than one idea (only 8.2% of times), if at all. The mean for QR7 was significantly lower than the sample mean with a mean difference of -0.26, 95% CI [-0.30 to -0.23], $t(1283) = -15.09, p < .000$. Regarding non-readers, QNR5 (M=0.76, SD=0.63) and QNR4 (M=0.88, SD=0.48) both scored lower than the other stems with an average number of ideas below 1 and significantly lower than the sample mean by -0.36 (95% CI, -0.41 to -0.31, $t(595) = -13.88, p < .000$) and -0.24 respectively (95% CI, -0.28 to -0.20, $t(595) = -12.33, p < .000$).

6.1.3 Length of User Responses. Table 3 presents the length of user responses for all sentences. QNR3 prompted the longest answers (52 characters on average) whereas QR6 was the shortest one (20 characters). The average length for the item assessing a *specific dimension of UX* QR6 was significantly lower than the sample mean (M=33) with a mean difference of -12.45, 95% CI [-14 to -10.9], $t(1283) = -15.7, p < .000$. Except for QNR2, all mean differences to the sample mean of 33 are statistically significant at $p < .000$.

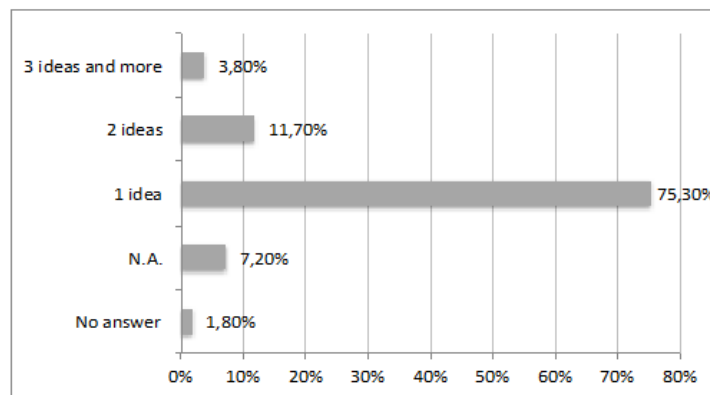
One logically observes significant correlations for all items between the length of responses and the number of ideas shared (from a minimum Person correlation coefficient of $r = .497$ to a maximum of $r = .769$ with all $p < .000$).

6.2 Qualitative Metrics

6.2.1 Variety of User Ideas. We categorized user ideas for each sentence stem using inductive coding (cf. Appendix A). We used the number of answer categories emerging from this analysis as an indicator of the richness of the outputs collected in terms of variety. As shown in Table 4, the number of categories ranged from 8 (QR6) to 17 (QNR3) depending on the sentence stem (sample mean 13.2). Sentence stems prompting users’ frustrations (QNR3, QR3, QR5) led to more variety, while both the overall experience item QR2 and the specific item QR6 collected a lower number of categories (9 and 8, respectively). Results of comparative stems are heterogeneous: 16 categories of insights were covered by non-readers on QNR1 but only 9 for readers on QR1. The *Extreme item* QR7 covered 13 answer categories, a variety score close to the mean of the sample. Interestingly, the two *redundant items* both covered 16 categories of answers related to user frustration, with an overlap on 12 of them. There were thus 4 additional answer categories elicited through QR5. We collected new insights in 62.9% of the cases using a redundant item and partially new ideas in 9.5% of cases (Table 5).

Table 2: Number of ideas produced per sentence stem (N=1880)

Sentence stem	Min	Max	Mean (SD)	No answer	0 (N.A)	1	2	≥ 3	Total ideas
QNR1	0	5	1.28 (0.64)	2 (0.3%)	/	466 (78.2%)	94 (15.8%)	34 (5.7%)	767
QNR2	0	4	1.14 (0.54)	6 (1%)	21 (3.5%)	482 (80.9%)	68 (11.4%)	19 (3.2%)	678
QNR3	0	7	1.39 (0.85)	2 (0.3%)	3 (0.5%)	435 (73%)	108 (18.1%)	48 (6.2%)	831
QNR4	0	3	0.88 (0.48)	4 (0.7%)	100 (16.8%)	464 (77.9%)	24 (4%)	4 (0.7%)	524
QNR5	0	4	0.76 (0.63)	22 (3.7%)	172 (28.9%)	361 (60.6%)	31 (5.2%)	10 (1.7%)	454
QNR6	0	8	1.29 (0.85)	10 (1.7%)	10 (1.7%)	459 (77%)	78 (13.1%)	39 (6.5%)	768
Total NR	0	8	1.12 (0.66)	1.3%	8.6%	74.6%	11.3%	4%	4022
QR1	0	5	1.32 (0.64)	1 (0.1%)	/	963 (75%)	246 (19.2%)	74 (5.8%)	1700
QR2	0	5	1.14 (0.44)	6 (0.5%)	/	1127 (87.8%)	127 (9.9%)	24 (1.9%)	1461
QR3	0	6	1.18 (0.71)	5 (0.4%)	95 (7.4%)	938 (73.1%)	180 (14%)	66 (5.1%)	1520
QR4	0	11	1.32 (0.74)	6 (0.5%)	10 (0.8%)	972 (75.7%)	204 (15.9%)	92 (7.2%)	1696
QR5	0	5	1.04 (0.64)	14 (1.1%)	177 (13.8%)	897 (69.9%)	160 (12.5%)	36 (2.8%)	1331
QR6	0	4	1.01 (0.47)	86 (6.7%)	13 (1%)	1074 (83.6%)	97 (7.6%)	14 (1.1%)	1313
QR7	0	6	0.86 (0.62)	94 (7.3%)	223 (17.4%)	862 (67.1%)	84 (6.5%)	21 (1.7%)	1100
Total R	0	11	1.12	2.4%	5.8%	76%	12.2%	3.7%	10121
TOTAL	0	11	1.12	1.8%	7.2%	75.3%	11.7%	3.8%	14143

**Figure 3: Distribution of the number of ideas provided by participants in response to each sentence stem (N=1880)****Table 3: Length of users' responses in number of characters (spaces included). (profile NR n=596; profile R n=1284)**

Sentence stem	Min	Max	Mean (SD)	Mean difference
QNR1	2	182	23 (22.9)	-9.97
QNR2	1	309	30 (33.6)	-2.29 (NS)
QNR3	1	335	52 (45.5)	19.39
QNR4	1	180	36 (28.1)	3.34
QNR5	1	197	27 (25.8)	-5.74
QNR6	1	378	24 (35.8)	-8.94
QR1	2	364	26 (30.4)	-6.54
QR2	1	320	23 (27.2)	-9.65
QR3	1	623	44 (43.1)	11.36
QR4	1	456	47 (42.4)	14.70
QR5	1	426	41 (37.1)	8.38
QR6	1	427	20 (28.4)	-12.45
QR7	1	409	38 (37.8)	5.44
Total	1	623	33 (33.7)	

Table 4: Variety and originality of user ideas are analyzed through the number of answer categories and the percent of stereotypical, unusual, and unique ideas.

item	answer categories	stereotypical ideas	unusual •unique ideas
QNR1	16	30 %	23.6 % •7.8 %
QNR2	16	16.4 %	24.9 % •2.8 %
QNR3	17	23.5 %	19.7 % •2.2 %
QNR4	12	23.6 %	13.9 % •5.5 %
QNR5	14	24 %	26 % •4.8 %
QNR6	10	32.1 %	14.7 % •8.1 %
QR1	9	40.1 %	9.8 % •4.8 %
QR2	9	59.8 %	7.7 % •2.2 %
QR3	16	18.1 %	17.4 % •2.2 %
QR4	11	44.6 %	20.8 % •6.9 %
QR5	16	27.6 %	16.3%* •3.8 %
QR6	8	49.6 %	13.8 % •1.8 %
QR7	13	22.8 %	22.7 % •5.8 %
Total	12.7	27 %	19.9 % •4.5 %

Table 5: Qualitative comparison of insights between the redundant items QR3 and QR5 (n=1284)

QR5 presents a:	Frequency	%	Valid %
Different idea	807	62.9%	73.6%
Partially different idea	122	9.5%	11.1%
Similar ideas	167	13%	15.2%
Missing data	188	14.6%	

Redundancy was thus worth in 72.4% of cases, especially since the qualitative analysis of the answers shows that these cover more hedonic aspects. An assumption here is that pragmatic issues would come to people’s minds and be mentioned first. In practice, it would be possible to aggregate the outputs to both questions in order to get an even more precise analysis in terms of the number of occurrences of each problem described by users.

6.2.2 Originality of User Ideas. To assess the originality of each sentence stem, we compared each idea provided by respondents to the frequency of its occurrence across participants. An idea mentioned by only 5% of respondents is considered ‘unusual’, and an idea mentioned by only 1% of respondents is deemed ‘unique’ [54]. We labeled as ‘stereotypical’ the answers that represented the top-1 answer category for each item.

On average, our sentences triggered 27% of stereotypical answers, 19.9% of unusual ideas, and 4.5% of unique ideas. The *generic item* QR2 produced a very high percentage of stereotypical data (59.8%) similar to the *specific item* QR6 (49.6%). On a qualitative side, we also observed that this *specific item* produced less elaborated ideas with numerous statements “it depends” or “should be enhanced” without much explanation. Several respondents mentioned that they did not understand the word “interface”. The *redundant item* QR3 produced few stereotypical data (18.1%), whereas its twin QR5 – presented second in the order of questions - resulted in more stereotypical data (27.6%). Once aggregated with the data of QR3, the percentage of unusual ideas for QR5 drops from 28.3% to 16.3% (marked with an asterisk in Table 4). However, some of the unusual

ideas collected are novel: the lack of social aspects in e-reading, as well as content quality issues and inspiring anecdotes about book covers, appeared only in QR5. The *extreme item* QR6 produced less stereotypical data (22.8%) and more unusual or unique ideas (22.7% and 4.5% respectively). Other stems scored higher than average on these two metrics, for instance QNR1 and QR4.

6.2.3 Analysis of the Valence. We analyzed the valence of users’ answers on four items (Table 6). We categorized as “neutral” the responses that do not fall clearly into either the positive or negative categories (for instance describing an experience as “ordinary” cannot be interpreted by the researcher as either positive, nor negative). Ambiguous responses that might go into either category depending on the interpretation one assigns were also categorized as neutral. We categorized as “mixed” answers encompassing both a positive and a negative element.

Respondents expressed a majority of unequivocal opinions, either positive or negative. When asked about their global experience (QR2), e-book readers gave more nuanced answers with 18.3% of neutral or mixed answers.

7 DISCUSSION

We conducted a two-level analysis of the results of this research. On the one hand, we provided accounts on qualitative data on domain-specific insights gained about the experience of e-reading. To understand the value of SCT to collect user research data, we discuss the commonalities and differences between these insights

Table 6: Analysis of the valence of participants' replies

Valence	Positive	Negative	Neutral	Mixed
QNR1	8.6%	82.7%	4.5%	4.2%
QNR6	96.7%	0.9%	2.3%	0.2%
QR1	77.4%	14.1%	3.7%	4.8%
QR2	63.9%	17.8%	12.5%	5.8%
QR6	64%	22.4%	7.1%	6.4%

and existing work on e-reading experiences. On the other hand, we provided a detailed formal analysis of the sentence stems themselves to explore how their construction could influence participants' answers and derive potential recommendations for the design of SCT studies.

7.1 The Experience of E-reading

The insights collected using SCT in our study show strong parallels with previous work on the experience of e-reading. At an overall level, the most adequate basis of comparison for our study is industry-based reports, featuring large-scale surveys (using stratified sampling strategies) on e-reading often accompanied or preceded by a qualitative stage. The yearly large-scale survey commissioned by the French publishers' association (SNE) presents very similar findings as ours [55, 56]. Data from a sample of N=502 e-books readers shows that digital books are preferred for ease of storage (70%), transportation and mobility (62%), pricing (59%), ease to acquire a new book (48%), and ease to pay (44%). On the other hand, paper books are preferred to offer or share (74%), pleasure and comfort (72% and 63%), quality and variety of choice (63% and 58%), and ease to annotate (54%). In a similar study [26], the sub-sample of e-reader respondents (n=1007) were asked about the main benefits they saw in e-books: 79% of occurrences refers to a pragmatic reason (e.g., storage, transportability, practical for the holidays, ubiquitous access to a personal library), 56% to an economical reason (some books are free in a digital format, which was the first advantage mentioned, and the fact that e-books are cheaper) and 52% for the ease to access e-books (fast, easy, access to books not easily available in paper format). On the drawbacks side, 68% of negative aspects mentioned relate to the experience of reading (screen fatigue, unpleasant to read outside, no sensoriality, hard to focus, lack of immersion). The second main family of factors frustrating e-readers is linked to ownership (no possibility to share them, we don't own the book). Our own results related to e-books readers are very similar (QR1), with transport and storage mentioned as the main advantages of e-books as compared to paper books (40%), followed by practical reasons (25%), pricing (6%), and accessibility (6%). A key difference is the prevalence of pricing in the industry surveys as compared to our findings, where the price is mentioned as an asset in both QR1 and QR4 but not as a dominant dimension. In QR3 and QR5, price is mentioned as a limitation and 12% of occurrences mention a lower price when people are asked what they dream of regarding e-books. This aligns with the contrasting picture depicted by industry surveys about e-reading practices (71% of e-readers in [26] read exclusively or almost exclusively free e-books). While the key findings of our

study are comparable to these industry surveys, all [26, 55, 56] used a mixed-methodology to complement the trends identified with qualitative insights (interviews or focus groups). Due to surveying major predefined dimensions only, these surveys also do not gather unusual or unique ideas from respondents, which is relevant when seeking trends but a limitation in a design context. Using the SCT, many respondents provided insightful explanations about their spontaneous replies, thereby giving us actionable and inspiring material to be used in the design process

A vast body of empirical studies on e-reading in academia focuses on students or library populations [11] or investigate the usability of a specific platform or e-reading device using user testing [7, 13, 16, 24, 43, 61, 72]. Regarding the benefits of e-books, the students surveyed by Höver et al. [24] (survey, N=62 students) see most benefits in portability, space-saving, and instant delivery. Features such as annotations or augmentation with multimedia content were also appreciated. Gibson and Forbes [16] (usability test, N=33, students) usability test included a question comparing the pros and cons of e-books over paper books for 4 types of electronic readers. All types confounded, advantages mentioned were storage, portability, and environmental benefits. The main disadvantages were limited battery life, potential damage to the e-reader, usability, cost, eye strain, and absence of sensorial elements. Storage and portability are similarly perceived as positive in our findings, and usability, cost, and absence of sensorial elements as disadvantages. However, we see that some dimensions have a different weight in the evaluation of the experience. For instance, the absence of sensorial aspects is way more dominant in our study than the one of Gibson and Forbes [16] and vice-versa for technology-related aspects such as battery life, which was mentioned by our participants (7% of occurrences on QR3, an additional 4% of QR5) but not as the top dimension as is the case for [16]'s computer science students sample. Besides the difference of technology literacy, user testing a device and answering many questions on its usability might orient the main assessment of the experience towards more practical or technology-related issues.

Our findings regarding hedonic and sensorial aspects align well with the notion of haptic dissonance by Gerlach and Buxmann [14] based on N=30 interviews conducted with readers. Haptic dissonance refers to the fact that e-books miss the well-known haptic elements of a paper book, and that the different "feel" of an e-book creates unpleasantness for the users. The 14 haptic aspects and underlying 16 dissonance statements identified by the authors in their qualitative study were all recognizable in our dataset. For instance, the resistance of our NR group in QNR1, with accompanying statements such as "it feels cold, impersonal", "I can't feel the progress

I have made” or QNR3 “I never read any e-books because it is meaningless, it does not equal the sensation of a book held in one’s hands and the way the paper feels when turning pages”. We find similar statements from e-book readers, especially in QR5 “what frustrates me the most with e-books”.

7.2 Benefits and Limitations of SCT in the Landscape of User Research Methods

It is useful to reflect on the nature of insights provided by previous studies on e-reading and how they compare to the data SCT (used as a remote survey method) provided in our use case. First, it seems like the main categories we identified (e.g., of issues, frustrations, or benefits) are similar to the ones identified in large-scale surveys using a qualitative stage (interviews or focus groups) to design relevant survey questions [26, 55, 56]. What these surveys lack is the rationale behind choices and verbatims, which they extract from the qualitative preliminary data. [26]’s survey for instance conducted four focus groups of 3h30 each (N=32) to this end. With a focus on the main dimensions, they also lack details on less common categories of content and their prevalence amongst specific types of users. Respondents also tend to agree with most of the statements presented (e.g. finding all features important on an e-book platform). As a benefit, quantitative surveys offer the clear benefit of being efficient when it comes to the analysis of the data. Conducted by specialized institutes, these specific surveys use stratified sampling strategies, samples are therefore representative of a specific population and the results and trends detected can be generalized. Similar to ours, the samples from other studies in the field are usually non-probabilistic. Our SCT items achieved a high item response rate (91% on average) and reached a relatively large sample of respondents (N=1880) in a short time. This is rather impressive for an online study disseminated on social networks with no incentives. Dissemination strategy and format might explain this success. First, we adopted a strategic dissemination strategy by reaching out online to large and engaged communities of readers (i.e. online reading clubs). Second, some participants and community leaders spontaneously shared the study link to their network, describing it as ‘fun’, ‘different than regular surveys’, and pleasant/easier to fill out. These reactions and the fact that people spontaneously shared the study tend to confirm the attractiveness of SCT as a survey format for respondents.

Our results are also partially comparable to standardized and theory-based survey studies on technology acceptance and e-reading, such as Tsai [63] (survey, N=213 book readers), Camarero et al. [5] (survey, N=227 users of e-readers), Gerlach and Buxmann [15] (survey, N=180, general population) or Dimitriadis et al. [11] (survey, N=441, general population). Indeed, these studies have complemented the traditional dimensions of the Technology Acceptance Model [10] with specific measures relevant for e-reading experiences such as price, compatibility with the preferred way of reading, and compatibility with past experiences. All these elements were highlighted as relevant by our sample, yet our SCT study does not aim to model the causal relationships between factors of acceptance. This comparison with standardized questionnaires yet raises the question of whether SCT studies should be based on a specific theory of experience. In psychology, several SCT tests

are based on specific theories, for instance the Sentence Completion Method by Rohde [47] based on Murray’s theory of needs. In the present study, and in most previous studies using SCT in user research, SCT does not explicitly or accurately rely on theories of user experience. The dimensions we chose to base our sentence stems on relate to generic user experience dimensions (such as expectations or needs) but were not purposively linked to a specific theory of experience. The idea is however interesting to explore: building SCT stems or analyzing data based on a specific UX theory. In fact, in the domain of psychology, the flexibility of the method has led to a proliferation of SCT forms, many of which were designed for a single study. As explained by Goldberg [18], “the ease of constructing sets of stems, the content of which bears a prima facie relationship to the variables under investigation, has encouraged a variety of research”. For HCI, suitable examples of theories for SCT could be the self-determination or psychological needs theories (based on [53]) which are already popular in UX research and which we used to ideate new e-reading experiences as a follow-up to the present study (anonymized). It can also suit any major model of user experience [21, 38] which are the basis of UX evaluation scales (Attrakdiff [22], UEQ [36], meCUE [41]). Our findings could be analyzed through the lens of the modular component model of User Experience [41], which inspired our dimensions, distinguishing product perceptions (divided into instrumental and non-instrumental), emotions, consequences (e.g. intention to use), and overall attractiveness.

Another common methodology adopted in the literature to study e-reading is usability testing of different e-books formats, platforms, or devices (e.g., [16, 61, 72]). These studies have a narrower focus and involve a relatively smaller sample size, yet they have the benefits of studying the experience in-situ. As compared to our study, findings from usability studies include precise identification of usability issues, cognitive measures, and actionable recommendations for the (re)design of the studied interfaces. Findings also tend to give more emphasis on technical and task-related aspects of the interaction. Our own results do include insights on some technical issues and usability concerns, along with interesting verbatims, as illustrated by this respondent’ reply to QR3 “The issue with e-books is. . . the difficulty to assess the length of a text and to navigate within the book, for instance finding an information on a page you previously read is particularly painful because there is no physical memory of its location as there is for a paper book”. Admittedly, verbatims with such length are the exception more than the rule in our SCT dataset and considerably depend on the sentence stem used. With the sample size we reached, we have a sufficient amount of detailed user inputs to be used for ideation. One strategy we used was to extract these detailed replies and to use them as raw inspiration material in an ideation workshop we conducted subsequently to the study (*anonymized*). Our study is however not suited to identify accessibility issues. While these were mentioned in several items, they were often minor in terms of the number of occurrences. As a reminder, 6% of our participants declared having a disability likely to impact their reading experience. A specific analysis of their replies might shed another light on these aspects.

Finally, the main comparable methodologies when it comes to the richness of insights gathered and their level of originality or uniqueness are qualitative in nature with a broader scope than

usability testing. The aforementioned interviews of Gerlach and Buxmann [14] are such an example: we indeed collected data showcasing all the 14 haptic dissonance aspects. As our investigation scope was larger, we also collected rich qualitative data on other dimensions. Gerlach and Buxmann’s sample was composed of 30 participants yet they reached saturation around $N=20$. We did not attempt to calculate which sample size in our case would have brought a saturation of insights, yet considering the qualitative data analysis efforts involved in an SCT study it would be worth considering which sample size would be “good enough” to understand the phenomenon studied. As highlighted by Swallow, Blythe & Wright, “qualitative data provides a richness and detail that may be absent from quantitative measures” ([60], p. 91), but collecting large amounts of qualitative data is a challenge. While replies to SCT are more straight to the point than interview data and already more clearly categorized by questions, they are comparable to open-ended questions in surveys (yet we hypothesize that the attractiveness of the format and response rate is higher), and the underlying analysis efforts should not be neglected. Previous SCT studies have often limited the data analysis to easier techniques such as word clouds or affinity diagramming [30, 31], which could also be suitable for Industry.

7.3 Sentence Stems Construction

Our in-depth analysis of sentence stems outputs highlighted relevant findings.

Looking at the influence of sentence stems on participants’ answers, we observed that our *generic sentence stem* QR2 produced around 60% of stereotypical ideas. While serving as a good onboarding process, the generic item in the present study was not the most insightful. *Comparative sentence stems*, having a very high item response rate, seemed easier to answer. This might be due to the fact that e-books have a real-world equivalent, which is often a basis for comparison [14]. Interestingly, the *redundant sentence stem* provided additional data (72.4% of new insights, and almost 20% of unusual or unique insights in response to the second item). Introducing redundancy was worth it, especially since the qualitative analysis of the answers shows that these cover more hedonic aspects. We could relate this finding by comparing redundant items to the 5 Whys or laddering technique [1], prompting people to provide more depth into their answer and often leading to a higher level of abstraction in the rationale of their experiences. One could also imagine sentences using two blanks to complete in order to collect both a feeling and the rationale behind it (e.g., When I read an e-book, I feel . . . because. . .). The lowest item response rates correspond to items prompting about inexistent expectations or aspirations, e.g. asking non-readers to imagine under which conditions they would read an e-book (QNR4) or what they expect from it (QNR5). Similarly, our *extreme sentence stem* asking users about their dream (QR7) also led to a lower response rate. It is however worth highlighting that this item collected ideas of a higher degree of novelty as compared to any other item. On the one hand, stereotypical answers (as the ones collected through the generic stem) can be considered good for emphasizing main issues or needs. However, on the other hand, variety indicates that the exploration of user insights has been thorough and will thus support the ideation

stage in the design process. In each survey, a good balance between both is thus desirable. It makes sense to introduce redundancy for items that produce few stereotypical data (as Q3, with only 18% of those replies), as they do not trigger consensus and one will not reach saturation quickly, it means that it is worth exploring them further. In UX design, one can stress the importance of originality in collected user insights: issues or needs that are rarely cited might be a better source of inspiration for designers. They might also reflect the fact that these original answers are more intimate, more personal, sometimes even shared with a story.

The length of answers also differed according to the sentence stem. The average length for the *specific item* QR6 was significantly lower than the sample mean, covering fewer answer categories and less elaborated ideas with rather frequent “it depends” or “to be enhanced” types of replies. This might be explained by an issue interpreting the items, some people in our sample being unfamiliar with the concept of ‘interface’. Of course, longer answers do not necessarily reflect more quality, yet it is worth pilot-testing the survey to make sure that the lack of clarity of an item will not impede the richness of the data.

It is unfortunately hard to compare our formal analysis findings to previous results from the literature in UX due to the fact that (1) very few studies in our field have applied the sentence completion method (2) these studies were limited in the way they analyzed participants’ responses, mainly presenting generic findings, often by counting the number of occurrences of insights and presenting them in a single table or word cloud [30, 31].

7.4 Limitations and Future Work

While we put a lot of effort to maximize the validity of our study, some limitations can be highlighted. First, it would have been insightful to compare a larger number of variations in sentence stems in a more experimental setup. While some previous studies in the UX literature entail more sentences to complete (such as [30] using 50 sentences or [31] using 24 sentences), they had comparatively way smaller sample sizes and often used incentives. With the objective of deriving guidelines for research and practice about the use of the SC method, we aimed for a realistic lightweight setup leveraged by a relatively large sample size (as compared to the aforementioned previous research using SCT and local standard in HCI [4]). We focused on a small number of variations only, representing typical variations used in user research studies. We also focused on a single case study, limiting the generalizability of our results.

Another limitation concerns the fact that the items were not randomized. As in traditional surveys, the order in which the questions are asked can have a significant influence on the results, as respondents might provide biased responses to subsequent items based on previously viewed items [37]. Based on our pilot tests and our experience using the SCT method in previous studies, we decided to maintain a logical sequence from the most general to the most specific question for all participants. In future studies, it might however be relevant to randomize questions to investigate a potential question-order effect. For instance, around a quarter of replies to question QR2 describe the user experience of e-books in comparison with paper books, being either described as less good

(9%), similar (7%), better (5%), or different (4%). Whether this is an intuitive way for people to describe e-books, or whether these replies have been prompted/biased by the preceding question QR1 being explicitly a comparison question cannot be shown in our study due to the absence of randomization. In the domain of personality assessment, Holaday et al. [23] recommend a random ordering of sentence stems, to “catch clients off guard” and collect more spontaneous and open responses. In UX studies, especially remote surveys, the social desirability stakes seem lower for respondents yet this is a worthwhile consideration for future work. Lastly, we also observed a limitation with regard to the survey branching. Participants who had already read an e-book once but declared not reading e-books at the start of the survey were presented with the sentence stems for non-readers. The item QNR3 (“I have never read any e-books because”) was thus too strict and led some respondents to simply reply that they had read an e-book already. In some cases, they commented on why they did not like it but it does not hold true for all replies.

Future work on the optimization of sentence stems could of course attempt to replicate our findings in different contexts, but also explore more variation in the type of stems. The position of the ‘blank to be filled’ in the sentence could be located at the beginning rather than at the end.

Finally, while it was not the objective of the present study, further studies are necessary to understand how the data obtained through the SCT compares with user data collected through other methods, such as user interviews for instance. How many interviews would one need to collect the same richness of data as gathered from 1880 respondents through SCT? And vice versa, how many respondents would one need for an online SCT study to reach the richness of data collected through user interviews? Is there a type of data that one could collect through one method and not the other? How does the effort of analyzing the data of SCT compare to open-ended questions or interviews? And if data collection is easier with SCT, is there a sample size over which the ratio of benefit vs. efforts to analyze is unbalanced? While a lot of research papers advocate for triangulation of data [45], not much has been published on how to do it efficiently in user studies.

8 RECOMMENDATIONS FOR THE DESIGN OF EFFECTIVE SENTENCE COMPLETION SURVEYS

Based on our analyses, we have derived a number of recommendations for researchers and practitioners willing to use the sentence completion method in their projects. Some of these recommendations are generally accepted good practices in any UX study, which we indicate for the sake of completeness. Note that our recommendations are based on a few items only (indicated below for each point) and our study did not rely on an experimental setup.

- As a general rule, first define what you want to learn about and brainstorm several sentence stems on each dimension under investigation. Use cognitive interviewing to pilot test sentence stems on a diversity of end-users in order to find the right wording (i.e. clear and unambiguous which triggers effectively the dimension under investigation) and to select the best stems.

- Prompting generic dimensions such as the overall experience is good as a scaffolding question. As an example of a generic dimension, we studied item QR2 “The reading experience on an e-book is”. In the present study, it did not lead to a wide variety of answer categories but an analysis of the valence of answers (positive, negative, neutral) offers a valid alternative to Likert scales.
- Comparative items (QNR1 and QR1 “Compared with a print book, an e-book is”) collected a high response rate and a high number of ideas. They seem interesting when there is an equivalent to the experience you seek to explore. Print books vs. e-books, e-voting vs. paper voting, a competitor’s application vs. yours, etc. Otherwise, one might formulate a sentence stem to elicit comparative experiences people can think of (“e-voting should be as secure as ___ e.g., online banking).
- Don’t be afraid to use redundant items to dig deeper into the dimensions that are of most interest to your project. Our item QR3 “The problem with e-books is” was followed by QR5 “What frustrates me the most with books”, both eliciting issues and frustrations. We collected new ideas in 63% of the cases using a redundant item (QR5) and also covered more hedonic aspects. Items that seem suited for duplication are the ones that produce less stereotypical data (thus requiring more user input to explore the topic).
- Using “extreme” items is a double-edged sword. In our study, item QR7 “I dream of an e-book that” led to a lower item response rate and a smaller number of ideas, but produced more original and novel users’ ideas. If you want to include this type of sentence stems, you might decide to make these items optional to not generate discomfort in respondents who would find them extravagant.

9 CONCLUSION

In this paper, we investigated how participants in an online survey respond to different types of sentence completion items (aka. sentence stems) in a large-scale study involving 1880 participants. By comparing our findings with previous work on e-reading experiences and formally analyzing the outputs of four types of stems on the quantity, variety, and originality of generated user insights, we confirmed the effectiveness of this method, whose outputs might compare to the rich insights collected through face-to-face interviews. We consider SCT as a promising alternative to open-ended questions (which usually have a low response rate) or multiple-choice questions (which do not bring insights into the ‘why’ of UX) at a time when UX practitioners are striving for lightweight user research methods [67]. Less time-consuming than an interview, yet more informing than a multiple-choice questions survey, SCT has multiple assets. Thanks to the present study, we believe that using remote online sentence surveys might contribute to a more cost-effective collection of insightful qualitative data.

We derived recommendations for the design of effective sentence completion surveys for both research and practice. With this paper, we also contribute to the consolidation of user research methods, which are often simply considered a means to an end (of designing and evaluating experiences). With a majority of innovative

UX methods being in early phases in their development [32, 67], studies geared towards a continuous improvement of the methods we employ should have a high priority on the design/HCI community research agendas. This long-term effort should ideally be distributed over the community, thereby allowing all fellow scientists to be equipped with the most valid and reliable tools to tackle the challenges of our field.

REFERENCES

- [1] Vero Vanden Abeele and Bieke Zaman. 2009. Laddering the User Experience! *User Experience Evaluation Methods in Product Development (UXEM'09)-Workshop*.
- [2] Margaret Ann. 1964. Use of the Gilmore Sentence Completion Test as a predictive instrument in relation to the academic achievement of certain high school students.
- [3] Kenneth S. Bordens and Bruce B. Abbott. 2011. *Research Design and Methods. A process approach (8th edition)*. McGraw-Hill, New York, NY.
- [4] Kelly Caine. 2016. Local standards for sample size at CHI. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery New York, NY, USA, 981–992, DOI:10.1145/2858036.2858498.
- [5] Carmen Camarero, Carmen Antón, and Javier Rodríguez. 2014. Technological and ethical antecedents of e-book piracy and price acceptance: Evidence from the Spanish case. *The Electronic Library* 32, 4: 542–566. DOI:10.1108/EL-11-2012-0149
- [6] Nathan T. Carter, Michael A. Daniels, and Michael J. Zickar. 2013. Projective testing: Historical foundations and uses for human resources management. *Human Resource Management Review* 23, 3: 205–218. DOI:10.1016/j.hrmr.2012.12.002
- [7] Tara Cataldo, Trey Shelton, Steven Carrico, and Cecilia Botero. 2014. How Users' Perceptions of E-Books Have Changed—Or Not: Comparing Parallel Survey Responses. *Proceedings of the Charleston Library Conference*, 304–310, DOI:10.5703/1288284315587.
- [8] Gilbert Cockton, Sari Kujala, Piia Nurkka, and Taneli Hölttä. 2009. Supporting worth mapping with sentence completion. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5727 LNCS, PART 2: 566–581. DOI:10.1007/978-3-642-03658-3_61
- [9] Rue L. Cromwell and Richard M. Lundy. 1954. Productivity of Clinical Hypotheses on a Sentence Completion Test. *Journal of Consulting Psychology* 18, 6: 421–424. DOI:10.1037/h0059972
- [10] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. 1989. User acceptance of computer technology: a comparison of two theoretical models. *Management Science* 35, 8: 982–1003.
- [11] Efstathios Demitriades, Eastern Macedonia, and Dimitrios Chatzoudes. 2019. Factors Affecting the Acceptance of Electronic Books by the Readers. *International Journal of Business and Economic Sciences Applied Research (IJESAR)* 12, 1: 7–19. DOI:10.25103/ijbesar.121.01
- [12] Hermann Z. Ebbinghaus. 1897. Über eine neue methode zur prüfung meistiger fähigkeiten und ihre anwendung bei schulkindern [About a new method for testing intellectual abilities]. *Psychologie und Physiologie der Sinnesorgane* 13: 401–457.
- [13] Chrysoula Gatsou, Anastasios Politis, and Dimitrios Zevgolis. 2015. Exploring users' experience with e-reading devices. *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, IEEE, 833–839, DOI:10.15439/2015F197.
- [14] Jin Gerlach and Peter Buxmann. 2011. Investigating the acceptance of electronic books - The impact of haptic dissonance on innovation adoption. *19th European Conference on Information Systems, ECIS 2011*.
- [15] Jin Gerlach and Peter Buxmann. 2013. Analyzing electronic book acceptance: A compatibility perspective. *Proceedings of the Annual Hawaii International Conference on System Sciences*, IEEE, 2813–2822, DOI:10.1109/HICSS.2013.94.
- [16] Chris Gibson and Forbes Gibb. 2011. An evaluation of second-generation ebook readers. *Electronic Library* 29, 3: 303–319. DOI:10.1108/02640471111141061
- [17] Maaïke S. Goelema, Renske De Bruijn, Sebastiaan Overeem, Els Møst, Reinder Haakma, and Panos Markopoulos. 2018. Conceptions of sleep experience: A layman perspective. *BMC Research Notes* 11(1): 1–5. DOI:10.1186/s13104-018-3584-2
- [18] Philip A. Goldberg. 1965. A Review of Sentence Completion Methods in Personality Assessment. *Journal of Projective Techniques & Personality Assessment* 29, 1: 12–45. DOI:10.1080/0091651X.1965.10120175
- [19] Joy Paul Guilford. 1967. Creativity: Yesterday, Today and Tomorrow. *The Journal of Creative Behavior* 1, 1: 3–14. DOI:10.1002/j.2162-6057.1967.tb00002.x
- [20] Steve Harrison and Deborah Tatar. 2011. On methods. *Interactions*, 10–11.
- [21] Marc Hassenzahl. 2003. The Thing and I: Understanding the Relationship Between User and Product. In M.A. Blythe, A.F. Monk, K. Overbeeke, and P.C. Wright, eds., *Funology: From Usability to Enjoyment*. Kluwer Academic Publishers.
- [22] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. 187–196. DOI:10.1007/978-3-322-80058-9_19
- [23] Margot Holaday, Debra A. Smith, and Alissa Sherry. 2000. Sentence completion tests: A review of the literature and results of a survey of members of the Society for Personality Assessment. *Journal of Personality Assessment* 74, 3: 371–383. DOI:10.1207/S15327752JPA7403_3
- [24] Kai Michael Hover and Max Mühlhauser. 2014. The creation and evaluation of lecture E-books. *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies, ICALT 2014*, IEEE, 647–649, DOI:10.1109/ICALT.2014.190.
- [25] Wayne D. Hoyer and Deborah J. MacInnis. *Consumer behavior (4th ed.)*. Houghton Mifflin, New York, NY.
- [26] IFOP. 2014. Etude des perceptions et usages du livre numérique - Etude qualitative.
- [27] Evangelos Karapanos, Jean Bernard Martens, and Marc Hassenzahl. 2012. Reconstructing experiences with iScale. *International Journal of Human Computer Studies* 70, 11: 849–865. DOI:10.1016/j.ijhcs.2012.06.004
- [28] Pekka Ketola and Virpi Roto. 2008. Exploring User Experience Measurement Needs. Proceedings of the 5th COST294-MAUSE International Workshop on Meaningful Measures Valid Useful User Experience Measurement VUUM, 23–26.
- [29] Sari Kujala and Piia Nurkka. 2009. Identifying user values for an activating game for children. *MindTrek 2009 - 13th International Academic MindTrek Conference: Everyday Life in the Ubiquitous Era*, Association for Computing Machinery New York, NY, USA, 98–105, DOI:10.1145/1621841.1621860.
- [30] Sari Kujala and Piia Nurkka. 2012. Sentence completion for evaluating symbolic meaning. *International Journal of Design* 6, 3: 15–25.
- [31] Sari Kujala, Tanja Walsh, Piia Nurkka, and Marian Crisan. 2014. Sentence completion for understanding users and evaluating user experience. *Interacting with Computers* 26, 3: 238–255. DOI:10.1093/iwc/iwt036
- [32] Carine Lallemand. 2015. Towards Consolidated Methods for the Design and Evaluation of User Experience.
- [33] Carine Lallemand and avec Guillaume Gronier. 2018. *Méthodes de design UX: 30 méthodes fondamentales pour concevoir des expériences optimales*. Eyrolles, Paris.
- [34] Carine Lallemand and Vincent Koenig. 2017. Lab Testing Beyond Usability: Challenges and Recommendations for Assessing User Experiences. *Journal of Usability Studies* 12, 3: 133–154.
- [35] Carine Lallemand and Vincent Koenig. 2017. "How could an intranet be like a friend to me?" - Why standardized UX scales don't always fit. *Proceedings of ACM ECCE conference (ECCE '17)*, Association for Computing Machinery New York, NY, USA, 9–16, DOI:10.1145/3121283.3121288.
- [36] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In A. Holzinger, ed., *Lecture Notes in Computer Science*. Springer-Verlag, Berlin Heidelberg, 63–76. DOI:10.1007/978-3-540-89350-9_6
- [37] Jonathan Lazar, Jintuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann Publishers, Cambridge, MA, USA.
- [38] Sascha Mahlke. 2007. User Experience of Interaction with Technical Systems. Retrieved from <http://opus.kobv.de/tuberlin/volltexte/2008/1783/>
- [39] Julian Meltzoff. 1951. The effect of mental set and item structure upon response to a projective test. *The Journal of Abnormal and Social Psychology* 46, 2: 177–189. DOI:10.1037/h0058624
- [40] Daphne Menheere, Mathias Funk, Erik van der Spek, Carine Lallemand, and Steven Vos. 2020. A Diary Study on the Exercise Intention-Behaviour Gap: Implications for the Design of Interactive Products. *Proceedings of DRS 2020 International Conference*, Design Research Society London, UK, DOI:10.21606/drs.2020.329.
- [41] Michael Minge, Manfred Thüring, Ingmar Wagner, and Carina V. Kuhr. 2016. The meCUE Questionnaire: A Modular Tool for Measuring User Experience. In M. Soares, C. Falcao, and T.Z. Ahram, eds., *Advances in Ergonomics Modeling, Usability & Special Populations. Proceedings of the 7th Applied Human Factors and Ergonomics Society Conference 2016*. Springer International Press, Switzerland, 115–128. DOI:10.1007/978-3-319-41685-4_11
- [42] Bernard I. Murstein, Robert M. Colon, Nicole Destexhe-de Leval, and Martine Hoof-Van Van Parys. 1972. Influence of Stimulus Properties of the Sentence-Completion-Method on Projection and Adjustment. *Journal of Personality Assessment* 36, 3: 241–247. DOI:10.1080/00223891.1972.10119751
- [43] Eunil Park, Jungyeon Sung, and Kwangsu Cho. 2015. Reading experiences influencing the acceptance of e-book devices. *Electronic Library* 33, 1: 120–135. DOI:10.1108/EL-05-2012-0045
- [44] Robert F. Peck and MC Guire Carson. 1959. Measuring Changes in Mental Health with the Sentence Completion Technique. *Psychological Reports* 5, 1: 151–160. DOI:10.2466/pr0.1959.5.g.151
- [45] Ingrid Petterson, Florian Lachner, Anna Katharina Frison, Andreas Rieni, and Andreas Butz. 2018. A bermuda triangle? - A review of method application and triangulation in user experience evaluation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery New York, NY, USA, DOI:10.1145/3173574.3174035.
- [46] Kenneth E. Rogers, John Bishop, and Robert C. Lane. 2003. Considerations for the Use of Sentence Completion Tests. *Journal of Contemporary Psychotherapy* 33, 3: 235–242. DOI:10.1023/A:1023977704349
- [47] Amanda R. Rohde. 1957. *The Sentence Completion Method: Its Diagnostic and Clinical Application to Mental Disorders*. The Ronald Press Company, New York,

- NY, USA.
- [48] Björn Rohles, Vincent Koenig, Antoine Fischbach, and Franck Amadieu. 2019. Experience matters: Bridging the gap between experience- and functionality-driven design in technology-enhanced learning. *Interaction Design and Architecture(s)* 42: 11–28.
- [49] Virpi Roto, Effie Law, Arnold Vermeeren, and Jettie Hoonhout. 2011. *User Experience White Paper: Bringing clarity to the concept of user experience*.
- [50] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1: 92–96. DOI:10.1080/10400419.2012.650092
- [51] Joseph M. Sacks. 1949. The relative effect upon projective responses of stimuli referring to the subject and of stimuli referring to other persons. *Journal of Consulting Psychology* 13, 1: 12–20. DOI:10.1037/h0055829
- [52] Jami J. Shah, Noe Vargas-Hernandez, and Steve M. Smith. 2003. Metrics for measuring ideation effectiveness. *Design Studies* 24, 2: 111–134. DOI:10.1016/S0142-694X(02)00034-0
- [53] Kennon M. Sheldon, Andrew J. Elliot, Youngmee Kim, and Tim Kasser. 2001. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology* 80, 2: 325–339. DOI:10.1037/0022-3514.80.2.325
- [54] Chaehan So and Jaewoo Joo. 2017. Does a Persona Improve Creativity? *The Design Journal* 20, 4: 459–475. DOI:10.1080/14606925.2017.1319672
- [55] SOFIA, SNE, and SGD. 2016. 6e baromètre sur les usages du livre numérique - Un marché stable qui confirme la pratique d'un usage mixte: livres numériques et imprimés coexistent.
- [56] Sofia, SNE, and SGDL. 2014. *Baromètre des usages du livre numérique - vague 4*.
- [57] Laurence C. Soley and Aaron Lee Smith. 2008. *Projective techniques for social science and business research*. Southshore Press, Milwaukee, WI, USA.
- [58] Mark W. Stephens. 1970. Stimulus Pull as a Determinant of Individual Differences in Sentence Completion Responses. *Journal of Projective Techniques and Personality Assessment* 34, 4: 332–339. DOI:10.1080/0091651X.1970.10380258
- [59] Erik Stolterman. 2008. The nature of design practice and implications for interaction design research. *International Journal of Design* 2, 1: 55–65.
- [60] David Swallow, Mark Blythe, and Peter Wright. 2005. Grounding Experience: Relating Theory and Method to Evaluate the User Experience of Smartphones. *Proceedings of the 2005 annual conference on European association of cognitive ergonomics (EACE '05)*, Association for Computing Machinery New York, NY, USA, 91–98.
- [61] Esta Tovstiadi, Natalia Tingle, and Gabrielle Wiersma. 2018. Academic E-book Usability from the Student's Perspective. *Evidence Based Library and Information Practice* 13, 4: 70–87. DOI:10.18438/ebliip29457
- [62] David K. Trites. 1956. Evaluation of assumptions underlying interpretation of sentence completion tests. *Journal of Consulting Psychology* 20, 1: 8. DOI:10.1037/h0044329
- [63] Wen Chia Tsai. 2012. A study of consumer behavioral intention to use e-books: The Technology Acceptance Model perspective. *Innovative Marketing* 8, 4: 55–66.
- [64] Wenn Chieh Tsai and Elise Van Den Hoven. 2018. Memory probes: Exploring retrospective user experience through traces of use on cherished objects. *International Journal of Design* 12, 3: 57–72.
- [65] Karen Turnbow and Richard H. Dana. 1981. The Effects of Stem Length and Directions on Sentence Completion Test Responses. *Journal of Personality Assessment* 45, 1: 27–32. DOI:10.1207/s15327752jpa4501_7
- [66] Heli Kristiina Väättäjä, Paula Savioja, Virpi Roto, Thomas Daniel Olsson, and Jari Tapio Varsaluoma. 2015. User experience goals as a guiding light in design and development? Early findings. *INTERACT 2015 Adjunct proceedings*, University of Bamberg Press, 521–527.
- [67] Arnold P.O.S. Vermeeren, Effie Lai Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: Current state and development needs. *NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction*, Association for Computing Machinery New York, NY, USA, 521–530. DOI:10.1145/1868914.1868973.
- [68] Tanja Walsh, Piia Nurkka, and Sari Kujala. 2010. Exploring Cross-Cultural UX Evaluation with the Remote Online Sentence Completion Method Cross-Cultural Design. *Proceedings of the Ninth International Workshop on Internationalisation of Products and Systems IWIPS 2010*, London, UK, 11–22.
- [69] Tanja Walsh, Piia Nurkka, Helen Petrie, and Jaana Olsson. 2013. The effect of language in answering qualitative questions in user experience evaluation web-surveys. *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, OzCHI 2013*, Association for Computing Machinery New York, NY, USA, 73–82. DOI:10.1145/2541016.2541049.
- [70] Tanja Walsh, Piia Nurkka, and Rod Walsh. 2010. Cultural differences in smartphone user experience evaluation. *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (MUM '10)*, Association for Computing Machinery New York, NY, USA, 1–9. DOI:10.1145/1899475.1899499.
- [71] Charles A. Weisgerber. 1969. *Psychological assessment of candidates for a religious order*. Loyola University Press, Chicago.
- [72] Yuanxiang Zeng, Xue Bai, Jie Xu, and Chuqi Gong Huan He. 2016. The Influence of E-book Format and Reading Device on Users' Reading Experience: A Case Study of Graduate Students. *Publishing Research Quarterly* 32, 4: 319–330. DOI:10.1007/s12109-016-9472-5

A APPENDIX

Qualitative Analysis per Sentence Stem

In this Appendix, we present the summary of answer categories for each of the 13 sentence stems of the survey along with the number of occurrences in each category.

Answer category	Occurrences	%	Answer category	Occurrences	%
(QNR1) Compared with a print book, an e-book is . . .					
Cold / no soul	232	30	Missing social elements	15	1.9
Without pleasure	96	12.5	Negative digital support	9	1.2
Lacking sensorial elements	84	10.9	Pleasant	8	1
Practical	70	9.1	Too expensive	6	0.8
Not practical	55	7.2	Different	6	0.8
Less good	45	5.9	Less expensive	4	0.5
Eye strain	34	4.4	-----		
Less aesthetical	34	4.4	Other answers	36	4.7
Less stimulating	30	3.9	Don't know	3	0.4
			Total sum of ideas collected	767	
(QNR2) In my opinion, e-books are targeted to...					
Travelers	111	16.4	People using it for work	20	2.9
Everybody	107	15.8	Disabled people	18	2.7
Young people / students	98	14.4	Occasional readers	12	1.7
Tech savvy people	85	12.5	People who can't afford	12	1.7
People who read a lot	45	6.6	Others / not me	8	1.2
People who do not have storage space	35	5.2	No one	8	1.2
Rushed people	25	3.7	-----		
People who don't like to read	24	3.5	Other answers	19	2.8
People who like to read	24	3.5	Don't know	27	4
			Total sum of ideas collected	678	
(QNR3) I have never read any e-books because . . .					
I miss sensorial and haptic elements	206	23.5	Cold / no soul	17	1.9
I prefer print books	181	20.6	I like libraries	16	1.8
Reading comfort is low	153	17.5	Too expensive	16	1.8
I don't want to try / I have no interest	108	12.3	No social aspects	10	1.1
I don't have the device	46	5.2	I don't read much	10	1.1
Print books are more practical	34	3.9	It is a blasphemy	8	0.9
I like to have nice bookshelves	25	2.8	-----		
Did not had the opportunity to try	19	2.1	Other answers	11	1.3
I did try already	18	2.1	Don't know	0	0
			Total sum of ideas collected	878	
(QNR4) I would read an e-book if . . .					
I have no choice / I am forced to	146	23.6	It is a specific format	17	2.8
Never / I am not interested	96	15.5	Reading comfort was high	15	2.4
I have the e-book or device	89	14.4	I had visual deficiencies	13	2.1
The print book does not exist	78	12.6	It had added value	7	1.1
I am travelling / commuting	42	6.8	-----		
I want to / am motivated to	38	6.1	Other answers	34	5.5
The business model changes	35	5.7	Don't know	8	1.3
			Total sum of ideas collected	618	
(QNR5) I expect an e-book to . . .					
Nothing / no expectations	139	24	Have specific features	24	4.2
Be less expensive / free	62	10.7	Not have technical issues	13	2.2
Be attractive / pleasant	56	9.7	Be short / quick to read	13	2.2
Easy to use	49	8.5	Not kill print books	11	1.9
Practical (portability, storage)	44	7.6	Not cause eye strain	10	1.7
As good as a print book	40	6.9	-----		
Offer an enhanced experience	27	4.7	Other answers	28	4.8
Disappear / stop existing	25	4.3	Don't know	37	6.4
			Total sum of ideas collected	578	

	(QNR6) When I read a print book, I feel. . .				
Good / Happy / Fulfilled	227	32.1	Alive	9	1.3
In immersion / absorbed	206	29.1	Nothing / normal	9	1.3
Relaxed / serene / zen	123	17.4	Negative feeling	6	0.8
Free	38	5.4	-----		
Myself	16	2.3	Other answers	52	7.3
I have learned new things	12	1.7	Don't know	10	1.4
			Total sum of ideas collected	708	
	(QR1) Compared with a print book, an e-book is. . .				
Portable / easy to carry / not bulky	601	40.1	Functional / accessible	95	6.3
Practical (no reason given)	380	25.3	Different / complement. to	29	1.9
Less pleasant / attractive	176	11.7	-----		
Less expensive / economical / free	100	6.7	Other answers	118	7.9
			Don't know	1	0.1
			Total sum of ideas collected	1500	
	(QR2) The reading experience on an e-book is...				
Pleasant / Comfortable	796	59.8	Something to try	15	1.1
Unpleasant / Awkward	134	10.1	To improve	4	0.3
Less good than a print book	124	9.3	-----		
Similar to a print book	95	7.1	Other answers	25	1.9
Better than a print book	70	5.3	Don't know	10	0.8
Different	58	4.4			
			Total sum of ideas collected	1331	
	(QR3) The problem with e-books is...				
Lack of sensorial / haptic elements	232	15.3	Limited offering	58	3.8
Technical restrictions (DRM)	179	11.8	Missing features	51	3.4
Price (too expensive)	178	11.7	Lack of soul / cold	43	2.8
Navigation related-issues	168	11	Can't store on bookshelf	36	2.4
Technical issues	138	9.1 7.7	We have too many	17	1.1
Digital format	117		Can't get an autograph	12	0.8
Battery life	106	7	-----		
Eye strain / fatigue	83	5.5 4.5	Other answers	21	1.4
Nothing / There is no problem	68		Don't know	13	0.9
			Total sum of ideas collected	1520	
	(QR4) What I love about e-books is...				
Portability / Ease to store	654	44.6	Adaptability	59	4
Ease to acquire	146	10	Reading comfort	48	2.2
Features (search, notes, bookmark, etc)	137	9.3	Diversity of choice	29	2
Practical elements	104	7.1	Usability	28	1.9
Technical aspects (backlight, battery life)	75	5.1	-----		
Price	70	4.8	Other answers	101	6.9
			Don't know	15	1
			Total sum of ideas collected	1466	

(QR5) What frustrates me the most with e-books is...

Lack of sensorial / haptic elements	401	27.6	Poor quality of the content	58	4
Nothing / No frustration	159	10.9	Battery	54	3.7
Navigation related-issues	121	8.3	Missing features	52	3.6
Price (too expensive)	92	6.3	Lack of reading comfort	34	2.3
Technical restrictions (DRM)	86	5.9	Absence of book cover	19	1.3
Missing social aspects	72	4.9	Technical issues	18	1.2
Limited offer	70	4.8	-----		
Digital format	69	4.7	Other answers	56	3.8
Absence of a physical bookshelf	62	4.3	Don't know	32	2.2

Total sum of ideas collected 1455

(QR6) I find that the interface of an e-book is...

Pleasant / well-designed	673	49.6	Similar to print books	30	2.2
Disappointing / Unpleasant	224	16.5	Less nice than print books	18	1.3
Practical / customizable	164	12.1	-----		
Can be improved	65	4.8	Other answers	24	1.8
It depends on the format and device	50	3.7	Don't know	109	8

Total sum of ideas collected 1357

(QR7) I dream of an e-book that. . .

Incorporates innovative features	316	22.8	More books available	33	2.4
Shows better technical performances	178	12.9	More aesthetic	23	1.7
Is less expensive	172	12.4	Doesn't confer eyestrain	13	0.9
Reproduces the feeling of a book	123	8.9	Enhanced quality	13	0.9
Remains the same	53	3.8	-----		
Is more practical	45	3.2	Other answers	56	4.0
Takes benefit of technology	44	3.2	Don't know / Don't dream	274	19.8
Enhanced navigation	36	2.6	about it		

Total sum of ideas collected 1385