

Calling for Flexibility and Inclusivity Alongside Rigour: A Quick Snapshot of Meta-Research via Systematic Review Synthesis

Katja Rogers^{1,*}

¹University of Amsterdam, Science Park 900, 1098 XH Amsterdam, Netherlands

Abstract

Human-computer interaction (HCI) as a research field is very diverse in methods, values, and foundational understandings of what constitutes knowledge. This key characteristic is crucial to consider when setting an agenda for meta-research in HCI. In this workshop paper, I provide a quick snapshot of the approaches to meta-research found in the synthesis methods of systematic reviews at CHI—based on a TOCHI paper that I am presenting as part of the CHI 2025 journal track. Reflecting on how this may represent disciplinary differences, this emphasizes the need for flexibility and epistemic inclusivity when shaping the future of meta-research in an interdisciplinary domain like HCI.

Keywords

synthesis, systematic reviews, interdisciplinary, rigour, methodological flexibility

1. Introduction and Background

Methodological rigour is paramount for trustworthy and robust research outcomes and a stable knowledge foundation. This is in broad strokes agreed upon by most academic fields. However, there are many differing views about what *constitutes* rigour and related concepts like robustness or trustworthiness across epistemic paradigms—which directly impacts methodological choices and criteria for “goodness or quality” [1, 2].

HCI as a field that sits at the blurry intersection of science, engineering, and design [3] is highly diverse in methodological approaches and epistemic values. This contributes greatly to its broad problem-solving potential, as well as its ability to leverage both internal *and* ecological validity and to contribute polished design artefacts and design knowledge. Systematic review methods originate in the medical field and so conventionally reflect a strict hierarchy of evidence types that positions randomized-controlled trials as the gold standard for primary research. Thus, for systematic reviews that attempt to comprehensively synthesize knowledge from research publications, HCI’s diversity of method and produced knowledge types is particularly challenging.

There are distinct tensions in HCI resulting in part from the multidisciplinary and interdisciplinary nature of our field [3, 4, 5]. With the influence and interconnection of different disciplines come sturdy differences in opinion about what makes knowledge valuable (aiming for knowledge that is replicable, transparent, authentic, robust, ...) and what makes methods for knowledge generation “correct”, or “appropriate”, or even whether any *can* be determined to be “correct” ones. Further, there are explorations of and calls for more replication and open science principles in HCI [6, 7, 8], yet not all of these principles are compatible in the strictest sense with all the epistemic paradigms in HCI; not all research aims to be replicable. There are also increasing attempts to explore methodological choices in HCI (e.g., [9, 10, 11]), but along with this also comes opposition to conversations about rigour which is sometimes viewed as “policing” and a potential trade-off to reflection [12]. Similarly, there

Meta-HCI '25: First Workshop on Meta-Research in HCI, April 26, 2025, Yokohama, Japan

*Corresponding author.

✉ k.s.rogers@uva.nl (K. Rogers)

🌐 <https://ajkt.github.io/> (K. Rogers)

🆔 0000-0002-5958-3576 (K. Rogers)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

is increasing pushback against long HCI papers, and reference lists of increasing length [13, 14], yet part of phenomenon likely also—at least in part—reflects methodological differences and domain-based expectations about reflection and engagement with the broader literature [15].

All of these tensions reflect broader rumblings across academia where methodological rigour is defined in very different ways by researchers from different schools of thought. Outside of HCI too, there are increasing pushes for pushes for rigour and the open science movement to define their goals and principles in a more flexible manner that works across different disciplines and epistemic traditions [1, 16]. Similarly, in the past, I have suggested that we need calls for rigour but also inclusivity in the context of knowledge synthesis methods in HCI [17]. Knowledge synthesis methods like systematic reviews are crucial to identifying and refining robust answers in a field, as well as creating new knowledge and shaping future research directions, and thus make it a key aspect to explore when considering meta-research. However, there are many different definitions, variants, guidelines, and approaches to systematic reviews and knowledge synthesis broadly [18]. Complicating this further, systematic reviews are often not reported in as much detail as might be desired [11].

To advance both HCI and meta-research in HCI, we need a broad, critical conversation about the frameworks we use for synthesis, and whether they are suited for the kind of primary research prevalent in HCI. Inversely, we need to explore more substantially what kind of knowledge we can draw from the kinds of evidence and outputs (e.g., design artefacts) that our field produces, and which methods are best suited to do so across disciplinary divides.

2. Snapshot of Systematic Reviews Synthesis Methods at CHI

For this quick look into systematic review synthesis methods in our field, I draw on the dataset from the TOCHI paper by Rogers et al. [11] that presents an umbrella review¹ that identified papers published as CHI full papers prior to June 2021 that state to be a systematic review or to follow a systematic approach. The goal of this paper was to explore **reporting quality** in CHI systematic reviews. This aspect of “how we report research” in the context of systematic reviews falls at the intersection of the categories of “what we publish” and “how we evaluate research” in Table 1 of this workshop’s Call for Papers [13]—but is also a necessary precursor to understanding and assessing “how we conduct research”. For this position paper, I briefly illustrate a snapshot of this dataset, focusing on the self-reported synthesis methods and the referenced frameworks, guidelines, and examples that shaped them.

Dataset: Method and Snapshot The overarching dataset resulted from a search on the ACM digital library (DL) for papers in the CHI proceedings mentioning “systematic review” together or separately, filtered to research articles, for subsequent screening to identify papers that report a review and claim a systematic process in some way. The papers’ references were checked via snowballing to search for more CHI systematic reviews not found by the ACM DL search, resulting in an additional paper. All potential papers went through eligibility screening by two researchers and using a third researcher as a tie-breaker. Overall, this search resulted in a corpus of 41 papers, which underwent data extraction and appraisal following the same process. For more details on the search and selection stages, please refer to the original paper [11].

As part of that review², the papers’ self-reported synthesis methods were summarized and categorized through an iterative open coding approach as belonging to one of three core review archetypes: “narrative review”, “categorical reporting”, or “meta-analysis”. Similarly, the reviews were examined for whether they had cited specific reporting guidelines as frameworks (and if so, which ones), or referred

¹Labelled in this way as it better reflects the diverse nature of the examined corpus papers compared to the term “review of reviews”. Typically, both of these review variants attempt to synthesize a common answer based on evidence from (systematic) reviews focused on the *same* question(s)—in contrast, here, the corpus reviews all focused on very distinct research questions, and the umbrella review focused on their reporting practices.

²The analysis and synthesis overall is rather more extensive and explored the extent to which review papers met reporting guidelines, specifically PRISMA [19, 20] (42 items) for quantitative-synthesis reviews and ENTREQ [21] (21 items) for qualitative-synthesis reviews. This is omitted for this position paper for scope.

Table 1

The CHI review papers stating a systematic approach were classified as narrative reviews, categorical reporting reviews, or meta-analyses. Categorical reporting was the most common (~71%); there were only two meta-analyses in this dataset.

Review Archetype	No. in CHI Corpus until and incl. 2021
categorical reporting	29
narrative summary	10
meta-analysis	2
	41

to other reviews as an example for their conduct and/or reporting when describing their method. I focus on these two aspects (essentially, two columns in the dataset) as a snapshot for this position paper.

Results The most common review archetype presented a *categorical reporting* approach, in which elements in the corpus papers were categorized in some way, and then reported as frequencies (e.g., as percentages [22] or counts [23]). These categorizations are created for example via quantitative content analysis (e.g., [24]), or via inductive thematic analysis and affinity diagramming (e.g., [25]) or coding more generally (e.g., “all papers were coded in terms of aim of the study, measures, results, measuring times, study designs, participants, game genre, duration of gameplay, and if provided, definition of enjoyment” [23], or “we distinguish several types based on the literature reviewed earlier” [7]).

This reporting of frequencies of course also involved narrative summaries of the results—when such descriptions were predominant throughout without more than scarce frequency reporting, the paper was classified as a *narrative summary*, which was the second-most common archetype (e.g., [26]). The third archetype—the *meta-analysis*—was rare, with only two papers in the corpus [27, 28]. These involve a statistical approach to summarizing effects found in interventional studies (ideally in randomized-controlled trials) to calculate an overall effect size. An overview of review archetypes in this dataset is provided in Table 1.

At the same time, a fair number of these review papers cite either quantitative-focused synthesis reporting checklists or none at all. When a reporting checklist was cited, it was usually PRISMA [19] (9/41 overall, or 9/15 of those citing a framework), which was developed for quantitatively focused systematic reviews (with an interventional research question) and meta-analyses. However, in HCI review papers, the checklists or guidelines seem mainly applied to the search and selection stage of the review, e.g., providing a PRISMA flowchart to illustrate this part of the process, and following a multi-stage selection process for screening potential papers based on outlined inclusion and exclusion criteria³—but not necessarily addressing PRISMA items in subsequent stages of the review. For example, very few review conducted critical appraisal to determine risk of bias or similar: Stowell et al. [27] employed the CASP quality assessment tool, Esterwood et al. [28] conducted robustness checks, and Brulé et al. [30] using custom quality metrics.

The other frameworks cited were Kitchenham [31]’s guidelines for reviews in Software Engineering, and QUORUM (for meta-analyses) [32]. Fourteen papers (either instead of, or in addition to citing frameworks) referenced other review papers as examples that informed their method and approach (e.g., Mekler et al. [23] reference that their use of QUORUM is preceded by Bargas-Avila and Hornbæk [33]). Another fifteen referenced neither frameworks, nor examples (e.g., [34]). Table 2 shows an overview of this aspect of the dataset.

³For example, Hirzle et al. [29] apply a first screening stage to the papers’ title, abstract, and reference sections, and assess the full text of the remaining papers in a second screening stage.

Table 2

Some CHI review papers made reference to frameworks/guidelines or prior (systematic) reviews as examples when describing how their method was aligned. Frameworks were referenced similarly often as example papers. When frameworks were referenced, the PRISMA was the most common.

Reference to Frameworks or Examples	No. in CHI Corpus until and incl. 2021
none cited	15
examples cited for methodology	14
frameworks cited for methodology	15
... PRISMA cited as framework	9
... Kitchenham [31] cited as framework	5
... QUORUM cited as framework	3
combinations of the above	5

3. Discussion

In summary, CHI review papers that employ (at least in parts) a systematic process (at least until and including 2021) most commonly apply a kind of categorical reporting approach to their synthesis, often in the form of qualitative coding approaches. This—especially when coupled with a lack of a critical appraisal stage, which was rare in this dataset, and a broad, non-interventional research question [11]—more closely matches a *scoping review* (or systematic mapping study or scoping study [31]) than what is expected in a conventional systematic review.

Yet the frameworks we reference when describing the method (PRISMA, QUORUM, and Kitchenham [31]’s guidelines) generally target conventional systematic reviews with a quantitative synthesis to answer an interventional research question. Moreover, these guidelines also include—and in QUORUM’s case specifically expect—statistical meta-analyses. Yet the CHI proceedings only rarely contain meta-analyses. These are mismatches that deserve more attention in future meta-research in HCI.

Which Reporting Guidelines Work for Interdisciplinary Synthesis? There are a number of guidelines that may be better suited to the kind of research found in HCI. For example, guidelines for scoping reviews (e.g., their protocols [35], their conduct [36], and their reporting [37]) would match the broader research questions and methods better.

Even when conducting quantitative systematic reviews, other guidelines that specifically target and include the synthesis of non-interventional studies, such as the NIRO-SR [38], may be better suited. Additionally, criticisms raised against PRISMA have been improved on in guidelines such as ROSES [39] in the environmental sciences—yet a search of the ACM DL (dated 14 March 2025) suggests neither of these two examples has yet to be employed by any HCI papers.

Further, given the strong prevalence of qualitative and mixed-method research in HCI, synthesis methods and guidelines for qualitative reviews and mixed-methods reviews (e.g., [40, 41]) may be more helpful than PRISMA. The ENTREQ [21] is a starting point for qualitative reviews of a specific kind, but it disregards the need to address methodological limitations and does not expect reflexivity even though this is highly important for many types if not all of qualitative research [42].

Finally, it is worth noting that none of guidelines work for more constructivist approaches or traditional narrative review types, that tend to eschew checklists in principle. Reviews of this nature are also valuable to the field. The commonly used guidelines also do not work for synthesizing knowledge found in specific design instances and iterations, and knowledge contributions resulting from interaction design research [43]. What synthesis methods and reviews would look like if designed to gather and develop knowledge across this kind of work remains an open question.

What Kinds of (Meta-)Research Does HCI Need? I strongly believe that the multi- and interdisciplinaryity of HCI is a strength, and that it something we should try to incentivize and facilitate. In practice, I believe this means we need the following:

- More randomized controlled trials to be able to conduct more meta-analyses to provide more robust answers to targeted research questions. This also requires more attention to the reporting of effect sizes.
- Use of more applicable guidelines, frameworks, and other resources for reviews of non-interventional studies and scoping/mapping reviews for better reporting quality in secondary research.
- Use of appropriate qualitative review method guidelines, frameworks, and other resources for synthesizing qualitative research for better reporting quality in secondary research. Depending on the epistemic paradigm at play, this may indeed involve rejecting checklists entirely, but still alternative conceptualizations to rigour will likely apply (e.g., to assess trustworthiness and/or the ability to become a catalyst for action [2]).
- More meta explorations and development of review synthesis methods for design artefacts and design research.
- Broader community awareness of the different types of knowledge created through different types of primary and secondary research, as well as epistemic inclusivity and flexibility with regards to their assessment. The guiding questions in Rogers et al. [11] presented as prompts for reflection when reporting systematic reviews may be a useful resource for starting this process.

Overall, this means the need to consider different conceptualizations of rigour and related concepts within HCI, and doing so in a way that is flexible, pragmatic, and mitigates epistemic exclusion and entrenchment.

Acknowledgments

Sincere thanks to my co-authors on the TOCHI paper, and many discussions over the years with fellow systematic-review-lutionaries like Katie Seaborn and others.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

- [1] C. N. Steltenpohl, H. Lustick, M. S. Meyer, L. E. Lee, S. M. Stegenga, L. Standiford Reyes, R. L. Renbarger, Rethinking transparency and rigor from a qualitative open science perspective, *Journal of Trial and Error* 4 (2023) 47–59. URL: <http://dx.doi.org/10.36850/mr7>. doi:10.36850/mr7.
- [2] Y. S. Lincoln, S. A. Lynham, E. G. Guba, *The SAGE Handbook of Qualitative Research*, 2011, pp. 97–128.
- [3] W. E. Mackay, A.-L. Fayard, HCI, natural science and design: a framework for triangulation across disciplines, in: *Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '97*, Association for Computing Machinery, New York, NY, USA, 1997, p. 223–234. URL: <https://doi.org/10.1145/263552.263612>. doi:10.1145/263552.263612.
- [4] A. F. Blackwell, HCI as an inter-discipline, in: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15*, Association for Computing Machinery, New York, NY, USA, 2015, p. 503–516. URL: <https://doi.org/10.1145/2702613.2732505>. doi:10.1145/2702613.2732505.
- [5] E. Blevis, E. Stolterman, The confluence of interaction design & design: from disciplinary to trans-disciplinary perspectives, in: *Undisciplined! Design Research Society Conference 2008*, Sheffield Hallam University Research Archive (SHURA), 2009. <https://shura.shu.ac.uk/529/> [Conference or Workshop Item].

- [6] F. Echtler, M. Häußler, Open source, open science, and the replication crisis in HCI, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI EA '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–8. URL: <https://doi.org/10.1145/3170427.3188395>. doi:10.1145/3170427.3188395.
- [7] K. Hornbæk, S. S. Sander, J. A. Bargas-Avila, J. Grue Simonsen, Is once enough? On the extent and content of replications in human-computer interaction, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 3523–3532. URL: <https://doi.org/10.1145/2556288.2557004>. doi:10.1145/2556288.2557004.
- [8] N. Ballou, V. R. Warriar, S. Deterding, Are you open? A content analysis of transparency and openness guidelines in HCI journals, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445584>. doi:10.1145/3411764.3445584.
- [9] R. Bowman, C. Nadal, K. Morrissey, A. Thieme, G. Doherty, Using thematic analysis in healthcare HCI at CHI: A scoping review, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: <https://doi.org/10.1145/3544548.3581203>. doi:10.1145/3544548.3581203.
- [10] N. G. Hughes, J. R. Flockton, P. Cairns, Growing together: An analysis of measurement transparency across 15 years of player motivation questionnaires, *International Journal of Human-Computer Studies* 169 (2023) 102940. URL: <http://dx.doi.org/10.1016/j.ijhcs.2022.102940>. doi:10.1016/j.ijhcs.2022.102940.
- [11] K. Rogers, T. Hirzle, S. Karaosmanoglu, P. T. Palomino, E. Durmanova, S. Isotani, L. E. Nacke, An umbrella review of reporting quality in CHI systematic reviews: Guiding questions and best practices for HCI, *ACM Trans. Comput.-Hum. Interact.* 31 (2024). URL: <https://doi.org/10.1145/3685266>. doi:10.1145/3685266.
- [12] R. B. Robinson, E. D. Mekler, Reflections towards more thoughtful engagement with literature reviews in HCI, in: Proceedings of the Halfway to the Future Symposium, HttF '24, Association for Computing Machinery, New York, NY, USA, 2024. URL: <https://doi.org/10.1145/3686169.3686183>. doi:10.1145/3686169.3686183.
- [13] J. Oppenlaender, S. Malacria, X. Fang, N. van Berkel, F. Chevalier, K. Yatani, S. Hosio, Meta-HCI: First workshop on meta-research in HCI, in: Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems, CHI EA '25, Association for Computing Machinery, New York, NY, USA, 2025. URL: <https://doi.org/10.1145/3706599.3706723>. doi:10.1145/3706599.3706723.
- [14] J. Oppenlaender, Past, present, and future of citation practices in HCI (2024). URL: <https://arxiv.org/abs/2405.16526>. doi:10.48550/ARXIV.2405.16526.
- [15] A. Crabtree, H is for human and how (not) to evaluate qualitative research in HCI (2024). URL: <https://arxiv.org/abs/2409.01302>. doi:10.48550/ARXIV.2409.01302.
- [16] M. Pownall, C. V. Talbot, L. Kilby, P. Branney, Opportunities, challenges and tensions: Open science through a lens of qualitative social psychology, *British Journal of Social Psychology* 62 (2023) 1581–1589. URL: <http://dx.doi.org/10.1111/bjso.12628>. doi:10.1111/bjso.12628.
- [17] K. Rogers, K. Seaborn, The systematic review-lution: A manifesto to promote rigour and inclusivity in research synthesis, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: <https://doi.org/10.1145/3544549.3582733>. doi:10.1145/3544549.3582733.
- [18] M. Krnic Martinic, D. Pieper, A. Glatt, L. Puljak, Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks, *BMC Medical Research Methodology* 19 (2019). URL: <http://dx.doi.org/10.1186/s12874-019-0855-0>. doi:10.1186/s12874-019-0855-0.
- [19] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart,

- J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ* (2021) n71. URL: <https://doi.org/10.1136/bmj.n71>. doi:10.1136/bmj.n71.
- [20] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews, *Bmj* 372 (2021).
- [21] A. Tong, K. Flemming, E. McInnes, S. Oliver, J. Craig, Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ, *BMC Medical Research Methodology* 12 (2012). URL: <https://doi.org/10.1186/1471-2288-12-181>. doi:10.1186/1471-2288-12-181.
- [22] J. Bergström, T.-S. Dalsgaard, J. Alexander, K. Hornbæk, How to evaluate object selection and manipulation in VR? Guidelines from 20 years of studies, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21*, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445193>. doi:10.1145/3411764.3445193.
- [23] E. D. Mekler, J. A. Bopp, A. N. Tuch, K. Opwis, A systematic review of quantitative studies on the enjoyment of digital entertainment games, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, Association for Computing Machinery, New York, NY, USA, 2014, p. 927–936. URL: <https://doi.org/10.1145/2556288.2557078>. doi:10.1145/2556288.2557078.
- [24] J. Himmelsbach, S. Schwarz, C. Gerdenitsch, B. Wais-Zechmann, J. Bobeth, M. Tscheligi, Do we care about diversity in human computer interaction: A comprehensive content analysis on diversity dimensions in research, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–16. URL: <https://doi.org/10.1145/3290605.3300720>. doi:10.1145/3290605.3300720.
- [25] C. Parker, M. Tomitsch, N. Davies, N. Valkanova, J. Kay, Foundations for designing public interactive displays that provide value to users, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–12. URL: <https://doi.org/10.1145/3313831.3376532>. doi:10.1145/3313831.3376532.
- [26] F. Brudy, C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmoose, N. Marquardt, Cross-device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–28. URL: <https://doi.org/10.1145/3290605.3300792>. doi:10.1145/3290605.3300792.
- [27] E. Stowell, M. C. Lyson, H. Saksono, R. C. Wurth, H. Jimison, M. Pavel, A. G. Parker, Designing and evaluating mHealth interventions for vulnerable populations: A systematic review, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–17. URL: <https://doi.org/10.1145/3173574.3173589>. doi:10.1145/3173574.3173589.
- [28] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, L. P. Robert, A meta-analysis of human personality and robot acceptance in human-robot interaction, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21*, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445542>. doi:10.1145/3411764.3445542.
- [29] T. Hirzle, M. Cordts, E. Rukzio, J. Gugenheimer, A. Bulling, A critical assessment of the use of SSQ as a measure of general discomfort in VR head-mounted displays, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21*, Association for Computing Machinery, New York, NY, USA, 2021. URL: <https://doi.org/10.1145/3411764.3445361>. doi:10.1145/3411764.3445361.
- [30] E. Brulé, B. J. Tomlinson, O. Metatla, C. Jouffrais, M. Serrano, Review of quantitative empirical evaluations of technology for people with visual impairments, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–14. URL: <https://doi.org/10.1145/3313831.3376749>. doi:10.1145/3313831.3376749.
- [31] B. Kitchenham, Guidelines for performing systematic literature reviews in software engineering,

Technical Report, Keele University and University of Durham, 2007. EBSE Technical Report EBSE-2007-01.

- [32] D. Moher, D. J. Cook, S. Eastwood, I. Olkin, D. Rennie, D. F. Stroup, Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement, *The Lancet* 354 (1999) 1896–1900. URL: [https://doi.org/10.1016/s0140-6736\(99\)04149-5](https://doi.org/10.1016/s0140-6736(99)04149-5). doi:10.1016/s0140-6736(99)04149-5.
- [33] J. A. Bargas-Avila, K. Hornbæk, Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, Association for Computing Machinery, New York, NY, USA, 2011, p. 2689–2698. URL: <https://doi.org/10.1145/1978942.1979336>. doi:10.1145/1978942.1979336.
- [34] K. Caine, Local standards for sample size at CHI, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, Association for Computing Machinery, New York, NY, USA, 2016, p. 981–992. URL: <https://doi.org/10.1145/2858036.2858498>. doi:10.1145/2858036.2858498.
- [35] M. D. Peters, C. Godfrey, P. McInerney, H. Khalil, P. Larsen, C. Marnie, D. Pollock, A. C. Tricco, Z. Munn, Best practice guidance and reporting items for the development of scoping review protocols, *JBIE Evidence Synthesis* 20 (2022) 953–968. URL: <http://dx.doi.org/10.11124/JBIES-21-00242>. doi:10.11124/jbies-21-00242.
- [36] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, *International Journal of Social Research Methodology* 8 (2005) 19–32. URL: <https://doi.org/10.1080/1364557032000119616>. doi:10.1080/1364557032000119616. arXiv:<https://doi.org/10.1080/1364557032000119616>.
- [37] A. C. Tricco, E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, S. Hempel, E. A. Akh, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M. G. Wilson, C. Garritty, S. Lewin, C. M. Godfrey, M. T. Macdonald, E. V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, O. Tunçalp, S. E. Straus, PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation, *Annals of Internal Medicine* 169 (2018) 467–473. URL: <http://dx.doi.org/10.7326/M18-0850>. doi:10.7326/m18-0850.
- [38] M. Topor, J. S. Pickering, A. Barbosa Mendes, D. V. M. Bishop, F. Büttner, M. M. Elsherif, T. R. Evans, E. L. Henderson, T. Kalandadze, F. T. Nitschke, J. P. C. Staaks, O. R. Van den Akker, S. K. Yeung, M. Zaneva, A. Lam, C. R. Madan, D. Moreau, A. O'Mahony, A. J. Parker, A. Riegelman, M. Testerman, S. J. Westwood, An integrative framework for planning and conducting non-intervention, reproducible, and open systematic reviews (NIRO-SR), *Meta-Psychology* 7 (2023). URL: <http://dx.doi.org/10.15626/MP.2021.2840>. doi:10.15626/mp.2021.2840.
- [39] N. R. Haddaway, B. Macura, P. Whaley, A. S. Pullin, ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps, *Environmental Evidence* 7 (2018). URL: <http://dx.doi.org/10.1186/s13750-018-0121-7>. doi:10.1186/s13750-018-0121-7.
- [40] E. Aromataris, Z. Munn, JBI manual for evidence synthesis, 2020. JBI, 2020. Available from <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-20-01>.
- [41] J. Thomas, A. Harden, Methods for the thematic synthesis of qualitative research in systematic reviews, *BMC Medical Research Methodology* 8 (2008). URL: <http://dx.doi.org/10.1186/1471-2288-8-45>. doi:10.1186/1471-2288-8-45.
- [42] T. May, B. Perry, et al., *Reflexivity and the practice of qualitative research*, volume 109, Sage Los Angeles, 2014.
- [43] M. Wiberg, E. Stolterman, What makes a prototype novel? A knowledge contribution concern for interaction design research, in: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, NordiCHI '14*, Association for Computing Machinery, New York, NY, USA, 2014, p. 531–540. URL: <https://doi.org/10.1145/2639189.2639487>. doi:10.1145/2639189.2639487.