Would Human-Robot Interaction Conferences Benefit From More Formal Reporting?
Evaluating a Novel Study Reporting Form

Patrick Holthaus[1], Alessandra Rossi[2], Snehesh Shrestha[3,8], Wing-Yue Geoffrey Louie[4], Aysegul Ucar[5],
Daniel Hernández García[6], Frank Förster[1], Antonio Andriella[7], and Shelly Bagchi[8]

Abstract— In an interdisciplinary and evolving research field like human-robot interaction, clear and precise results reporting is essential for study comparability and replicability. To address the lack of a standard for such reporting and, at the same time, provide guidance for novices in the field, we have developed a web-based reporting form to capture human-robot interaction studies, serving as a model for how conferences could adopt it into the submission pipeline. In this work, we present a formative evaluation of this form regarding its level of detail, format and clarity, and the perceived benefits for authors, reviewers, and the community as a whole. We report the expert review of nine researchers who highlight the substantial value of this tool. In addition, these experts also provide suggestions for improvements to its form and the addition of details surrounding qualitative reporting.

I. INTRODUCTION

The research field of human-robot interaction (HRI) consists of a multi-disciplinary effort to understand and develop the relationship between humans and robots [1]. As a result, the different fields involved apply a vast range of methods to evaluate robot design or behaviours. While such innate diversity allows for invaluable insights from various perspectives and a vast pool of expertise, communicating between disciplines is often difficult because of differences in terminologies and reporting styles [2]. In contrast to other disciplines (e.g., Psychology [3] or Healthcare [4]), no standard for reporting has been established for the HRI community yet [5]. As a consequence, experimental outcomes are sometimes difficult to set into context or compare across different best practices. Moreover, the lack of details and

information, and the way these are presented in research papers, does not always allow the community to replicate and reproduce the work, leading to inconsistency and incongruence in the results across different research efforts [6].

There have been multiple attempts to formalise individual aspects of HRI experimentation - suggesting study design and methods [7], [8], metrics and reporting [9], [10], [6] - as well as efforts to create more comprehensive guidelines [11]. However, a consensus has not been reached as more nuanced and encompassing approaches are being suggested [12], [13].

Joining these efforts, a working group has been formed to establish a "Recommended Practice for Human Studies in Human-Robot Interaction" (IEEE P3108)[1]. In the coming year, IEEE P3108 is expected to publish a set of guidelines for HRI researchers for designing and conducting studies involving human participants. As part of this initiative, the group suggested a supplementary appendix [6] with the goal that conferences adopt it as a supplement for paper submissions. This postulation for standardised reporting as an appendix for conference papers constitutes a first step to address the issue of reproducibility and replicability in HRI [2], [14], and provides methodological structure to researchers while aiming to minimise additional work for conference authors, reviewers, and editors.

In this work, we present 1) the implementation of the appendix suggested in [6] as a new online reporting form, and 2) its formative evaluation relying on experts' opinions, whom we have surveyed to answer our research questions as follows:

RQ1 Are the questions in the reporting form appropriately detailed?

RQ2 Is the current format of the online reporting form clear and easy to understand?

RQ3 Which members of the HRI community can benefit from the reporting form?

In Section II, we review related work that highlights the need for standardised reporting to facilitate reproducibility and discusses the challenges that such attempts face. Section III describes how we derived an online study reporting form from the initially suggested appendix and its evaluation regarding the research questions above. Results of this study are presented in Section IV, and discussed in terms of insights and limitations in Section V. We conclude with a

[1]Robotics Research Group, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK. Email: {p.holthaus, f.foerster}@herts.ac.uk

[2] University of Naples Federico II, via Claudio 21, 80125, Naples, Italy Email: alessandra.rossi@unina.it

[3] University of Maryland, College Park, Maryland, USA. Email: snehesh@umd.edu

[4] Intelligent Robotics Lab, Oakland University, Rochester, Michigan, USA. Email: louie@oakland.edu

[5] Mechatronics Engineering Department, Firat University, Elazig, Turkiye. Email: agulucar@firat.edu.tr

[6] School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK Email: d.hernandez_garcia@hw.ac.uk

[7] Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028, Barcelona, Spain. Email: aandriella@iri.upc.edu

[8] U.S. National Institute of Standards and Technology, Gaithersburg, Maryland, USA. Email: snehesh.shrestha@nist.gov, shelly.bagchi@nist.gov

[1]https://sagroups.ieee.org/3108/

summary of findings and proposed future work in improving the online form according to the feedback gathered.

II. Related Work

In the past decade, the HRI community recognized a need for reproducibility and repeatability of user studies [6] by discussing guidelines for designing and reporting user studies. Several attempts have been started to discuss the challenges and requirements of creating common standards through workshops, tutorials, and papers (e.g., [15], [14], [16]). To improve upon the standardization of HRI experiments, numerous workshops have been facilitated at conferences, such as *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS) [15], *ACM/IEEE International Conference on Human-Robot Interaction* (HRI) [17], [18], [10], and the *IEEE International Conference on Human-Robot Interactive Communication* (RO-MAN)[2],[3] [19]. These workshops have often focused on highlighting the challenges with standardization in HRI experiments [17], facilitating discussions and brainstorming [17], disseminating state-of-the-art evaluation methods [20], or focusing on specific HRI application areas (e.g., human-robot teaming and accessibility) [21], [18]. Although these workshops have been fruitful for brainstorming, highlighting individual efforts, dissemination of the challenges, and motivating the importance of HRI standardization, their purpose was not to provide a formal agreed-upon set of standards for reporting for HRI studies involving human participants; so there remains a gap.

Similarly, a few research papers have drawn attention to the importance of reproducibility in HRI for advancing the field and making research more effective [19]. Likewise, a 2022 paper highlights the lack of reporting standards in HRI studies [5]. Their analysis of major conferences in the field reveals significant gaps in reporting.

Another paper [14] discusses broader challenges related to reproducibility in HRI, emphasising the field's interdisciplinary nature and the need for benchmarking, standardised methodologies, and transparent reporting. Their paper highlights the role of biases in research evaluation and suggests community-driven solutions for improving reproducibility. In the same vein, the paper [22] reviewed 18 conferences that each published at least 7 HRI papers in 2022. Their analysis found that only two of those conferences reported guidelines for promoting transparency and reproducibility. The authors highlight the difficulty in providing generalizable guidelines as well as clear recommendations as to how these should be applied within HRI.

While the above recommendations provide a strong foundation for improving transparency, implementing these guidelines requires a collective community effort and possibly conference mandates to ensure compliance. We introduced the first initial draft of the IEEE P3108 recommended practices appendix for study reporting in [6] to fill this gap in study documentation. However, the challenge to balance

TABLE I
PROPOSED STUDY REPORTING FORM BREAKDOWN

Pg.	Category	Fields
1	General information	Title, submitter name and email
2	Author information	Author names and affiliations
3	Study information	Location, date, setting, approval
4	Participants/recruitment	Number and justification, recruitment and target population, compensation, consent, demographics, criteria
5	Demographics	Categories, justification, statistics
6	Domain expertise	Categories, justification, statistics
7	Variables	Independent, dependent, control
8	Layout	Within, between, mixed
9	Hypotheses	Support, tests, preregistration
10	Execution	Overview, instructions, priming, factors, timing, exit, appearance, deception, debrief
11	Code/data	Code, datasets, repositories, scripts

comprehensiveness with ease of use remains, as overly detailed reporting requirements or a lack of clarity may deter researchers from full adoption.

III. Methods

While the form, which [6] proposed as a paper's appendix, provides a valuable foundation, this work presents a first step in investigating its practical use among experts in the HRI community. To evaluate its use, we first detail how we derive the online reporting form from the proposed appendix. To evaluate the resulting form against the research questions detailed above, we then present a study that has been approved by the University of Hertfordshire's Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (ECDA) under protocol number SPECS/SF/UH/05828.

A. Online Reporting Form Design

For ease of use and consistency, we developed the online reporting form using the *Gravity Forms*[4] plugin for the *WordPress* framework[5], and made it publicly available[6] for the purpose of this evaluation. We used full-screen width options for most fields during the digitisation process. To optimize space utilization and to reduce the page length, we used half page width for short fields and fields that qualified the primary fields. This includes names and email addresses, affiliated institutions, research locations, and their dates. Each field consists of a clear title, a sub-title describing the purpose of the field, and examples for further clarification. For multiple-choice fields, an *Other* option is always provided.

The form is divided into sections as per appendix [6] and represented as pages in this digital form as outlined in Table I. The order of most of the fields followed the original appendix, however, some fields were moved to group information based on their nature. For example, in the first section, the form now collects the title and author information

TABLE II

Survey questionnaire presented to participants via Microsoft Forms, open questions to "elaborate" on answers omitted.

#	Question	Type	Section
1	Consent	Yes/No	A
2	**Additional** important study **details** should be included in the study reporting form.	7-point scale	B
4	The study reporting form includes **too many details**.	7-point scale	
6	**The format** of the study reporting form makes it easy to fill.	7-point scale	C
8	The questions on the study reporting form are **clear and easy to understand**.	7-point scale	
	It would be **beneficial for** [...] if the study reporting form was a mandatory requirement when submitting a paper reporting about a study involving human human subjects to the HRI conference or equivalent venues.		D
10	[...] = **reviewers**	7-point scale	
12	[...] = **authors**	7-point scale	
14	[...] = **the HRI community**	7-point scale	
16	Please add any other **comments** you might have	Open	
17	If you would like to be contacted for a **follow-up semi-structured interview (using your publicly available corresponding email address associated with this paper)**, please check the following option.	Yes/No	E

before the study sections (i.e., location, dates, ethical review body name and approval numbers). We also consider that users may use multiple groups and combinations for items such as demographics, expertise, variables, study groups, hypotheses, and predictions. In this case, a plus button allows them to add additional rows for each field. This enables researchers to customize their input to the number of conditions and categories of participants' attributes that they considered.

The form allows the users to save the form at any point and receive an email notification of the form, so they can save and continue at any later time to allow for flexibility in filling the form. The form was iteratively validated internally with a pilot study with the P3018 standards appendix subgroup, where $N = 6$ participated by filling out the form and providing feedback until all members were fully satisfied and no additional issues remained.

B. Participant selection

In the study, we approached authors who conducted human subjects studies and published high-quality papers in peer-reviewed venues at two of the major HRI conferences, i.e., the *ACM/IEEE International Conference on Human-Robot Interaction* and the *IEEE International Conference on Human-Robot Interaction Communication*. We recruited from the last three years (2023-2025) and selected for high paper quality by only considering papers that were nominated or won best paper awards. Moreover, we filtered these papers to only consider those that present the results of a study with human participants. This allows participants to refer to their study design and thus helps us gather their feedback and opinions about the appendix from people who conducted a relevant experiment. In total, we compiled a list of 42 papers for recruiting participants with the goal of recruiting one author per paper.

From the pool of these papers, we invited authors to participate in this study via the corresponding authors' email addresses publicly displayed in the paper. If a corresponding author was unreachable or not available, we reached out to their co-authors in order of appearance and when email addresses were associated with the paper.

Whenever an author agreed to participate, we invited them to (i) read the paper introducing the initial appendix draft [6], (ii) review and fill in the appendix web form as detailed in Section III-A, (iii) fill in a short questionnaire using Microsoft Forms (see Table II), and (iv) optionally participate in a semi-structured interview to review and confirm if the intended meaning and the participant interpretation matched.

C. Preparation

We designed our survey so that responses to our questionnaire would be based on experience filling in the proposed study reporting form. We asked participants to first familiarise themselves with the nature of the form by reading the short paper by Bagchi et al. initially proposing it [6].

As a second preparatory measure, we asked participants to access and test the study reporting form online[3]. Participants were asked to provide data for the items outlined in Table I using either data from the publication they were contacted about or fake data if they preferred, e.g. for privacy reasons. The entered data were not evaluated, and any personal data entered were screened out since they are out of the scope of this study. Participants were offered the opportunity to delete any information after the questionnaire was completed.

D. Questionnaire

When participants had finished the preparation phase, by filling out the study reporting form, we arranged an online meeting to fill in a questionnaire. During this meeting, the experimenters introduced themselves, shared a link to the questionnaire and provided a brief overview of its structure. The experimenter then stayed in the background to be available for questions without influencing the participant while they filled out a questionnaire.

The questionnaire, as outlined in Table II, contained five sections, each presented on a separate page, covering (A) consent; (B) questions addressing the level of detail in the reporting form (**RQ1**); (C) questions addressing the reporting form format and clarity (**RQ2**); (D) questions addressing potential beneficiaries from adopting the reporting form (**RQ3**); as well as (E) general comments and follow-up;
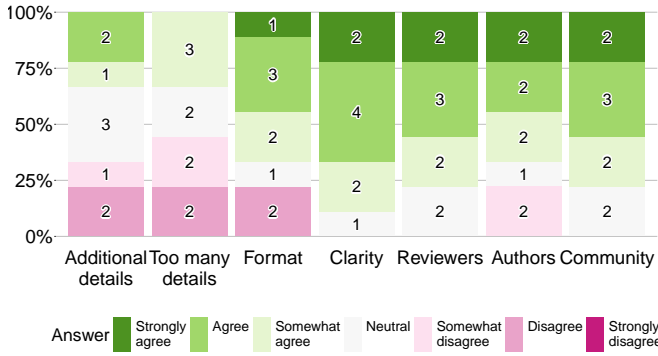
Fig. 1. Participant responses to scale-based questions about Additional details (Q2), Too many details (Q4), Format (Q6), Clarity (Q8), Benefits for Reviewers (Q10), Authors (Q12), and the Community (Q14) as described in Table II.

totalling 17 questions. We asked participants to rate the reporting form fields based on the appropriateness, clarity, and level of benefit for different members of the community using 7-point scales, labelled as *Strongly agree*, *Agree*, *Somewhat agree*, *Neutral*, *Somewhat disagree*, *Disagree*, and *Strongly Disagree*. Each of these questions was complemented with an open question asking to "please elaborate" to gain additional insights into participants' choices and considerations (questions 3, 5, 7, 9, 11, 13, and 15).

No audiovisual recordings were made, and we did not ask for personal information in any of the questions. Therefore, demographic information about age, gender, or occupation has not been collected. However, there was a potential for participants to share personal information, such as when reporting study results or giving examples. Whenever participants used personal information, the experimenter would notify them and anonymise the transcript if the participant requested so.

IV. RESULTS

In total, nine authors of high-quality papers familiar with study design techniques consented to participate in our initial evaluation and filled out the questionnaire. This section presents a brief descriptive statistic of the participant rating questions and results from the open questions to elaborate on participants' answers grouped by their thematic focus. As described in Section III-D, lower numbers in scale-based question ratings indicate higher agreement with a statement. To support the reporting of descriptive statistics and augment qualitative reporting, we provide a breakdown of responses to scale-based questions in Figure 1.

A. Study reporting form details

With regard to **RQ1**, some participants would like to see more details in the form, as indicated by their response to question 2 (see Table II), which were evenly distributed around the neutral value (4) with three ratings below and three above neutral (median 4), cf. Figure 1. This result was reinforced by answers to the open question 3, where participants requested options to provide information about group

interactions and structures, fields allowing better descriptions of qualitative studies, more rigorous questioning about study types, and adherence to data collection legislation (e.g., General Data Protection Regulation (GDPR)[7]).

At the same time, a few participants also suggested that some details could be omitted from the form. Responses for question 4 about too many details (median 4) also showed that three participants somewhat agreed that the form was overly detailed, whereas two participants were neutral, and the remaining four participants at least slightly disagreed with this statement (see Fig. 1). Comments to question 5 reveal that some questions might either not apply to a particular study but also that while there are lots of details to provide, these are mostly required and that it might be easier to fill in at submission time than in retrospect.

B. Study reporting form format

Concerning **RQ2**, the format of the reporting form has been rated as mostly positive, with a median value below neutral (3) for question number 6. Six of the participants found that the study reporting form is of appropriate format, whereas two others disagreed and one participant was neutral about whether the format makes the form easy to fill (see Fig. 1). Common concerns regarding the format were the size of text boxes, which should be increased, especially for items like hypotheses. However, the clear structure and examples were praised as useful in filling out the form.

The clarity of the form was generally seen positively and was rated as easy to understand, with all ratings of question 8, except one neutral, being better than average (median 2). Improvements around the addition of more information or examples were suggested, and the description of how to provide data about participant groups and hypotheses should be improved.

C. Study reporting form beneficiaries

About **RQ3**, participants found the form to be beneficial for all groups, including reviewers, authors, and the HRI community as a whole. Figure 1 shows that Questions 10, 12, and 14 have all been agreed upon (medians: reviewers 2; authors 3; community 2). The reporting form was rated as highly beneficial for reviewers, with all ratings apart from two neutral at least somewhat agreeing (Q10). Two participants disagreed somewhat with the usefulness of the suggested reporting form for authors, and one was neutral about it, while the other participants agreed (Q12). Two participants were neutral about the benefits for the HRI community while all seven other participants agreed with that statement to various degrees (Q14).

Participants commented (Q11) that the form provides additional helpful information for reviewers and that "*this study reporting form can help the reviewers to quickly understand the whole human subjects study process, which will improve work efficiency for reviewers.*" The reporting form

was also considered helpful, with space constraints imposed by conferences and could help cross-check experiments. Participants also noted that the data should ideally also be contained in the paper and otherwise requested by reviewers.

Comments about the benefit of the reporting form for authors (Q13) were mostly positive, stating that it might help to compact the paper submission itself but also allow "*to self-review their papers*" for any missing data or processes, which could be especially beneficial for early-career researchers to learn about human-subject studies. However, the additional workload was also commented on as potentially hindering while not providing essential benefits.

Participants commented about the benefits for the HRI community in question 15. They stated that the reporting form would "*absolutely benefit the community*" and can provide a "*standardize view to look into the papers, and their quality [sic]*." It was further mentioned that it can help to collect and analyse robotics studies related to human subjects and to share them with each other. Another benefit for the community was to make papers easier to fully understand the presented papers and to compare HRI studies. Caution was being urged not to hinder submissions with making this a mandatory requirement. Instead, the reporting form "*should be highly encouraged in the community*."

D. General comments

Participants provided some general comments as answers to question 16 that helped to augment the insights about the research questions and put them into perspective. Participants commended the opportunity to participate in this opportunity and stated that it was "*a great initiative and much needed*." Comments mentioned that the reporting form can help to standardise research practices but also "*serve as a guide especially for early researchers for conducting solid user studies*." However, the strong focus on quantitative studies was criticised, and additional elements were suggested to capture more qualitative studies.

V. Discussion and Limitations

Our initial evaluation of expert opinions has revealed that the proposed study reporting form is received generally positively and our findings suggest that such a form might contribute to the maturing of the field of HRI by establishing more rigorous reporting practices and providing guidelines for upcoming researchers.

With regard to the form's level of detail (**RQ1**), participants were generally satisfied with the questions' details as presented in the form, and no question was deemed inappropriate or impossible to answer. However, we also received mixed feedback where some parts were requested to be added to the form and others removed. On the one hand, including additional questions could particularly help to better capture qualitative approaches. On the other hand, some details were seen to be not applicable or unnecessary for certain study types. A potential solution would be to add conditional elements that are enabled depending on the type of study conducted.

Although we recruited experts in the field of HRI, their expertise was unlikely to cover all the potential expertise required to account for all types of HRI studies, especially those with unique populations, niche domain areas, or require special considerations, such as interaction length. In the future, it would be beneficial to further evaluate the required amount of details for the form with a population sample that provides a greater diversity of expertise for different HRI study types (e.g., longitudinal and special populations).

Apart from providing some longer descriptions and further examples, our findings attest that the reporting form is clear and easy to understand for experts in the field (**RQ2**). Nonetheless, the format as an online tool could be improved to allow for more flexible inputs, for example, by providing larger input fields for some of the questions. Follow-up research is required to assess whether the form is as understandable for novices reporting about their first HRI studies and whether accessibility mechanisms can be provided for this tool.

With the exception of one, none of the participants had difficulty with completing the online reporting form, which demonstrates that it was relatively easy to complete. This one participant experienced confusion regarding the email requirement to save and continue, mistakenly assuming that providing the authors' email addresses was sufficient. This technical limitation should be addressed in a future version of the reporting form.

However, it should be noted that the participants completing these forms are experts in HRI as the recruitment criteria were for individuals with awards from the conferences. In the future, it would be beneficial to get a broader participant pool of expertise in HRI to investigate how easily and quickly can the form be completed. This can also alleviate some of the participants' concerns regarding the potential workload of having to complete the form as a mandatory requirement. In particular, including early-stage researchers as participants could provide insights into whether offering more guidance, such as short explanations and examples, might enhance the report's comprehensiveness and usability.

To answer **RQ3**, we gathered feedback on groups that we thought might benefit in different ways from the proposed reporting form. Results show that authors, reviewers, and the community as a whole might benefit in different ways and some groups might benefit more than others. Common considerations regarding benefits were contrasting an increased workload that might hinder the acceptance of our reporting form with long-term benefits for all members of the community.

While the increased *effort* specifically for authors was highlighted repeatedly, the reporting form might decrease the mental load for reviewers. Higher author effort was also contrasted with providing a useful tool to check one's submission for incomplete reporting of data and procedures. Novice HRI researchers may especially benefit from such a reporting form as it can serve as not only a checklist for reporting but also support initial study design formulation as it highlights the core criteria for an HRI study design. It

was further mentioned that the workload would not increase drastically at submission time since the study design would still be in people's minds, unlike in the presented study, which looked at studies in retrospect.

Community benefits that were attributed to the reporting form, such as increased study replicability, the potential for a more standardised evaluation of studies, and facilitating information sharing, would highly depend on the community's willingness to adopt such practices. For that, we suggest a multi-step approach where an improved form could be tested as an optional supplement at conferences that are willing to spearhead our efforts in making HRI research stronger. Moreover, efforts from our side need to be undertaken to make sure that the form asks for appropriate details covering as many study types as possible, and that it is accessible to all members of the community without introducing too much overhead.

It should be noted that this study has some limitations. Firstly, the sample size was small due to the limited target population of authors of high-quality papers at established conferences in the HRI community. While this limits the generalisability of the results, it still provides highly valuable qualitative insights from a few select experts in the domain based on the design of studies that have actually been carried out, analysed, reviewed, and published. Secondly, the recruitment of participants primarily focused on experts in the field with high-quality papers that resulted in awards. It would be important to conduct further research on evaluating this form with an increased sample diversity to also include the entire community of HRI researchers, ranging from novices to experts, as they are the target audience of the reporting form. Thirdly, the current study focused on expert participants' experience filling in the proposed study reporting form. We expect that the adoption of the study reporting form by conferences and publications in the field of HRI will be beneficial for not only authors but also reviewers and researchers in the community as a whole. Additional evaluation is required where participants instead of filling the form, evaluate how detailed the information of an already filled form is, and how it helps them to better understand a research paper in the context of reviewing or replicating/reproducing its results. Finally, the benefits of the study reporting form have only been evaluated through subjective responses. As the central goal of the reporting form is to improve the reproducibility of HRI studies, it would be important to investigate the effects of the reporting form on reproducibility of research. This could be accomplished by improving the current version of the reporting form with the feedback gained in this study and then evaluating whether the introduction of the form leads to research that can be more consistently reproduced than when the form is not utilized.

VI. Conclusion

In this paper, we have presented the formative evaluation of a new study reporting form that has been proposed as a standardised appendix for conferences to adopt into their submission process [6]. Expert researchers' opinions about the form that has been adopted into an online format attested to this endeavour a great potential to benefit the HRI community, providing value for authors, reviewers, and readers alike. However, some suggested improvements to the reporting form demand a slightly improved format as well as more flexible input that can better adapt to study types that are common in the field, such as qualitative work. We aim to implement the feedback into an improved version of the form and evaluate the results in future work. With the hope of making the final reporting form a part of the reviewing process in HRI, this work will contribute to the maturing of the field, facilitating standardised reporting and providing guidelines and best practices for early-career researchers.

Disclaimer

This publication is intended to capture external perspectives related to standards, measurement, and testing-related efforts by the U.S. National Institute of Standards and Technology (NIST). This report was prepared in collaboration with the external authors; it is intended to document external perspectives, and does not represent official NIST positions. Further, certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available.

References

[1] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Šabanović, *Human-robot interaction: An introduction*. Cambridge University Press, 2024.

[2] B. Leichtmann, V. Nitsch, and M. Mara, "Crisis ahead? Why human-robot interaction user studies may have replicability problems and directions for improvement," *Frontiers in Robotics and AI*, vol. 9, 2022.

[3] A. E. Kazak, "Editorial: Journal article reporting standards." *American Psychologist*, vol. 73, no. 1, 2018.

[4] I. Simera, D. Moher, J. Hoey, K. F. Schulz, and D. G. Altman, "A catalogue of reporting guidelines for health research," *European journal of clinical investigation*, vol. 40, no. 1, pp. 35–53, 2010.

[5] J. R. Cordero, T. R. Groechel, and M. J. Matarić, "A Review and Recommendations on Reporting Recruitment and Compensation Information in HRI Research Papers," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 1627–1633.

[6] S. Bagchi, P. Holthaus, G. Beraldo, E. Senf, D. Hernández García, Z. Han, S. K. Jayaraman, A. Rossi, C. Esterwood, A. Andriella, and P. Pridham, "Towards Improved Replicability of Human Studies in Human-Robot Interaction: Recommendations for Formalized Reporting," in *International Conference on Human-Robot Interaction (HRI 2023)*. Stockholm, Sweden: ACM/IEEE, 2023, pp. 629–633.

[7] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme, "From characterising three years of HRI to methodology and reporting recommendations," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 391–398.

[8] K. L. Koay, D. S. Syrdal, K. Dautenhahn, and M. L. Walters, "A narrative approach to human-robot interaction prototyping for companion robots," *Paladyn, Journal of Behavioral Robotics*, vol. 11, no. 1, pp. 66–85, 2020.

[9] M. Zimmerman, S. Bagchi, J. Marvel, and V. Nguyen, "An Analysis of Metrics and Methods in Research from Human-Robot Interaction Conferences, 2015-2021," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 644–648.

[10] M. Zimmerman, A. Virts, S. Bagchi, S. Shrestha, P. Holthaus, E. Senft, D. Hernández García, and J. A. Marvel, "The Road to Reliable Robots: Interpretable, Accessible, and Reproducible Human-Robot Interaction (HRI) Research," in *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '25. IEEE Press, 2025, p. 1997–1999.

[11] G. Hoffman and X. Zhao, "A primer for conducting experiments in human–robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, pp. 1–31, 2020.

[12] K. Fischer, "Effect Confirmed, Patient Dead: A Commentary on Hoffman & Zhao's Primer for Conducting Experiments in HRI," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, pp. 1–4, 2021.

[13] J. Seibt, C. Vestergaard, and M. F. Damholdt, "The Complexity of Human Social Interactions Calls for Mixed Methods in HRI: Comment on "A Primer for Conducting Experiments in Human-robot Interaction," by G. Hoffman and X. Zhao," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, pp. 1–4, 2021.

[14] H. Gunes, F. Broz, C. S. Crawford, A. R.-v. der Pütten, M. Strait, and L. Riek, "Reproducibility in Human-Robot Interaction: Furthering the Science of HRI," *Current Robotics Reports*, vol. 3, no. 4, pp. 281–292, Dec. 2022.

[15] M. E. G. Dalgaard, M. Giuliani, T. Haidegger, A. Tapus, and G. S. Virk, "Workshop: Towards Standardized Experiments in HRI," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[16] U. Tijani, H. Wang, and Z. Han, "Towards Reproducible Language-Based HRI Experiments: Open-Sourcing a Generalized Experiment Project," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 949–953.

[17] D. Chrysostomou, P. Barattini, J. Kildal, Y. Wang, J. Fo, K. Daut-enhahn, F. Ferland, A. Tapus, and G. S. Virk, "ReHRI'17-Towards Reproducible HRI Experiments: Scientific Endeavors, Benchmarking and Standardization," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 421–422.

[18] J. A. Marvel, S. Bagchi, M. Zimmerman, M. Aksu, B. Antonishek, Y. Wang, R. Mead, T. Fong, and H. Ben Amor, "Test Methods and Metrics for Effective HRI in Real-World Human-Robot Teams," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 652–653.

[19] C. Jost, B. L. Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, and N. Mirnig, Eds., *Human-Robot Interaction: Evaluation Methods and Their Standardization*, ser. Springer Series on Bio- and Neurosystems. Springer Cham, 2020, vol. 1.

[20] C. Jost, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, B. Le-Pévédic, and N. Mirnig, "EMSHRI 2017: Third International Workshop on Evaluation Methods Standardization for Human-Robot Interaction," in *IEEE/RSJ IEEE International Symposium on Robot and Human Interactive Communication*, 2017.

[21] J. Marvel, S. Bagchi, M. Zimmerman, M. Aksu, B. Antonishek, J. Case, T. Fong, V. Nguyen, A. Norton, E. Senft, and Y. Wang, "Workshop on Test Methods and Metrics for Accessible HRI," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 944–946.

[22] M. Spitale, R. Stower, M. T. Parreira, E. Yadollahi, I. Leite, and H. Gunes, "HRI Wasn't Built In a Day: A Call To Action For Responsible HRI Research," in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 2024, pp. 696–702.