Letter to the Editors of *JUX*

Rolf Molich
DialogDesign
Skovkrogen 3
DK-3660 Stenlose
Denmark

I appreciate that JUX is publishing papers by promising young UX professionals, such as "Talking About Thinking Aloud: Perspectives from Interactive Think-Aloud Practitioners," by Liam O'Brien and Stephanie Wilson, in Volume 18, Issue 3, of the *Journal of User Experience*.

The authors of this article are clear about what they did; for example, they interviewed nine recent-entry UX professionals in one country. The authors also provide many thought-provoking quotes from the nine usability professionals that they interviewed.

But a number of issues in this article trouble me.

Generalization Value
JUX's call for papers states: "Authors are invited to submit manuscripts addressing various aspects of quantitative and qualitative usability studies that have a strong generalization value to other practitioners working with any human-interactive product."

The article bases its conclusions on a small convenience sample with just nine UX professionals from the UK. In light of this, I would argue that this paper does not have a "strong generalization value."

No Interrater Reliability
The interviews and most of the analysis were carried out by the first author only. Interrater reliability is neither discussed nor available. The authors should have addressed this limitation clearly in the abstract, the conclusion, and the section "Limitations."

Discussion Guide
The article is based on nine interviews. The article mentions a Discussion Guide, but few details are provided. Because a good discussion guide is essential for a valid interview, such details would have provided a better understanding of the validity of the results. Besides, I have not been able to find information about the length of the interviews. Despite several requests to the authors, I have not received the Discussion Guide. The authors should have put the Discussion Guide in an appendix to the article.

What Is ITA?

The distinction between Traditional Think-Aloud (TTA) and Interactive Think-Aloud (ITA) is unclear and seems poorly defined in the article. In TTA, "the moderator silently observes the test session, except for issuing occasional reminders to 'keep talking,'" whereas in ITA, "the moderator makes interventions while the participant is using the system, such as asking questions." However, some of the examples of ITA intervention types in Tables 2 and 3 are used by all usability moderators. For example, practical types of intervention in TTA include situations in which a user has misunderstood the task or encounters a technical problem.

The article quotes Goodman et al.'s (2012) definition of usability tests as "structured interviews focused on specific features in an interface prototype." I agree that the use of ITA in structured interviews is helpful, but I disagree that a structured interview is a usability test.

We may have a naming issue here. We should not call all interactions in which a user works with a prototype or product a "usability test." Different types of interactions should be considered different techniques, each with its own name.

The authors should have reported the participants' views of the key differences between usability testing, semi-structured interviews, and design critiques.

Types of ITA
Table 2 lists 13 "Intervention types: Data usefulness" whereas Table 3 lists 6 "Intervention types: Practical." Some of these intervention types are always applicable, whereas others may be acceptable if used responsively, and some may distort the results. This distinction is not made in the article.

The authors should have concluded with a discussion of "levels of ITA" and the types of ITA that are appropriate for each method: usability testing, semi-structured interview, and design critique.

Is ITA a Good Way to Do Things?
The article does not address whether what people are doing in ITA is actually a good way to do things. It is quite possible that, although many people practice ITA, what they are doing may be problematic—and they do not even realize the biases and problems that they are causing by their moderating style. Usability studies are robust in the sense that, if you make a variety of mistakes, you may still uncover valid problems. So, it is quite possible that much of what happens in ITA is not the best approach, but because the practitioners following it still often obtain useful results, they do not realize that what they are doing is causing problems and that there are better ways to moderate.

Opinions
My clients demand evidence, not opinions. To me, the main strength of TTA is "shut up and observe" rather than ITA's "stop and probe." This difference is crucial because TTA avoids contamination of the data by test participants' opinions and unintended hints from the moderator. Results from TTA are free from opinions. Opinions can be dangerous in organizations with a low UX maturity because skeptics can argue, "Why are participants' opinions better than mine?" Skeptics cannot use this argument against low effectiveness and efficiency observed or measured using TTA. In addition, TTA is simple to teach and practice. With ITA, usability testing loses its distinctive character and becomes just another opinion-based method.

The authors should have reported the participants' experiences reporting opinions resulting from ITA.

Conclusions

The article contains a number of conclusions, for example:

- Participants often believed they could compensate for validity problems in their analysis and reporting or through triangulation with other methods.
- [Participants] did not think that usability testing could provide this kind of valid behavioral data anyway.

- This flexibility can even extend to the repurposing of the usability testing session from a primarily observational study to something more like a semi-structured interview or a participatory usability evaluation.

It is not clear from the article how many of the nine participants agreed or disagreed with each of these interesting conclusions.

Tips
The "Tips for User Experience Practitioners" section at the end of the article is helpful.

I would have expected the tips to be based on the conclusion, but that does not seem to be the case. The authors should have explained the relation between "Tips for User Experience Practitioners" and the rest of the article.

Talking with Participants Is Not Usability Testing
Just to be clear, I do not think there is anything wrong with talking with a participant, asking their opinions about things and why they are doing what they are doing, or even how they might want to design the system. You may get some insights while doing this, but this is not the way to run a usability test session. My problem with the attitude of many of the participants in this research is that they do not seem to understand that, although anything is permissible, there are good reasons for running a session one way and not another. There is nothing wrong with talking to strangers in an elevator about problems they have with some system, but this is not a usability test.

JUX's Review Process

Acceptance
If I had been a reviewer, and if the authors had not addressed my concerns, I would have recommended rejecting the paper.

My original letter to the Editors criticized both the paper and the review process. The Editors responded, "The letter should not be addressed to our reviewers or about our review process, but only focused on the paper itself." The lack of guidance from the three reviewers and the manuscript manager to the authors of the *JUX* paper raises serious doubts about *JUX*'s review process, error culture, and the quality of published papers. It should be a prominent goal for the Editors and reviewers to constantly strive to improve their skills and the quality of *JUX*.

Rolf Molich

DialogDesign

About the Author

Rolf Molich
Mr. Molich manages DialogDesign, a small Danish usability consultancy. He conceived and coordinated the Comparative Usability Evaluation (CUE) studies where more than 140 usability professionals evaluated the same applications. He is the co-inventor of the heuristic inspection method (with Jakob Nielsen). He received the UXPA Lifetime Achievement Award in 2014.