Research on quality in interpreting

We all want a quality performance. But how do you define that?

Elisabet TISELIUS.
Published: October 29, 2012 Last updated: December 2, 2015

Jérôme, one of the 2interpreters, Michelle (Interpreter Diaries) and myself have been involved in a discussion on how to evaluate interpreter exams - really tricky business as anyone who has been on an exam jury will know. Jérôme published a really interesting reflection on final exams, and Michelle and I responded. You can read the post here.

We have now arrived at the even trickier subject of quality in interpreting and this is where I felt I needed to write a post, not just continue with comments. Clearly what exam jurors are after is some type of high quality interpreting, and this is also supposedly what accreditation jurors or peer-assessors are looking for. But what is it?

Determining quality

Michelle mentions two early studies, one by Hildegund Bühler (questionnaire study with interpreters as respondents) and the other by Ingrid Kurz (questionnaire study with interpreting users as respondents). These two have recently been followed by another with interpreters as respondents by Cornelia Zwischenberger. When talking about questionnaire studies, it should also be mentioned that AIIC commissioned a study made by Peter Moser on user expectations, and that SCIC regularly make surveys of their users’ expectations.

Bühler and Kurz more or less conclude that interpreting is good when it serves its purposes and that different contexts have different requirements (I’m summing up really heavily here).

As both Michelle and Jérôme point out in their comments, there is a flood of articles on quality, and many studies, but I’m not sure we have actually come up with something more conclusive than Bühler and Kurz did. However, I would like to draw your attention to something that I have found most interesting in research on quality.

Barbara Moser-Mercer was also mentioned in the comments and she published an article in 2009 challenging the use of surveys for determining quality. This seems inspired by the work that has been done in Spain by Angela Collados Ais and her research team ECIS in Granada. She only publishes in Spanish and German, so I had to travel there to understand what she does. It was worth it - extremely interesting research. I also have to compliment them on how I was received as a guest. Emilia Iglesias-Fernandez made me feel like royalty, and all the other researchers in the unit were extremely welcoming and accommodating.

But here’s the interesting thing: For the past 10 years they have been researching how users of
interpretation perceive and understand the categories most commonly used in surveys to assess interpreting. These categories have typically been, since Bühler: Native accent, pleasant voice, fluency of delivery, logical cohesion, consistency, completeness, correct grammar, correct terminology, appropriate style. If I remember correctly, for instance, the AIIC Survey on expectations of users of conference interpretation showed that experienced users considered content particularly important and cared more about correct terminology and fluency than pleasant voice or native accent.

In their experiments the ECIS team has been tweaking interpreted speeches so that the exact same speech would be done with or without native accent, with or without intonation, high speed or low speed and so forth. Different user groups first rated how important the different categories were and then they were asked to rate different speeches, tweaked for certain features. When you do that it turns out that the exact same speech with native accent gets higher score for quality (e.g. using more correct terminology or correct grammar) than the speech with non-native accent. And the same goes for intonation, speed and so forth.

So to make a point, I would posit that features that are not rated as important (such as accent) affect how the user perceives important features (e.g. content or correct terminology).

In interpreting research there is also a lot of error analysis going on of course, and many studies base their evaluation of the interpretation on error analysis. One problem with that is exactly the one that Jérôme points out – maybe the interpretation actually got better because of something that the researcher/assessor perceived as an error. Omissions, for example, are especially difficult to judge in that regard. I have also gotten results with my holistic scales where the interpreter that I perceived as “much better” (only gut feeling) got much worse scores. One reason for this could very well be that that interpreter omitted more, and thereby, in comparison with the source text, there were more “holes” or “faults” or whatever you would like to call them.

What about exams?

When it comes to exams, Jérôme claims that not much has been done in terms of research on exam assessment. I have not checked that, but my impression is that he is right. I cannot remember reading about quality assessment of examinees. Both Daniel Gile and Franz Pöchhacker point to the work of Emma Soler-Caamaño (again, unfortunately for me, in Spanish). It looks like she’s unique when it comes to looking at exams. Gile, Pöchhacker and Sawyer all seem to say that (sadly) not much research has been done on interpreter education (although Sawyer has indeed made an extensive effort to remedy that). I know that aptitude tests and entrance exams have received attention from researchers, but final exams… Please enlighten me.

Jérôme also brings up the question of training of exam jurors or peer-reviewers. This is one of my pet subjects, but on which there seems to be very little consensus, at least in the environments I’ve worked in. Now, I don’t mean to say that there are no courses on how to be an interpreting exam juror; of course there are. But what I mean (and Jérôme too I think) is that people evaluating interpreting do not get together and discuss what they believe is good interpreting or not. One could, for instance, organize a training session before an exam to bring jurors together to discuss criteria and how they understand them, and also to listen to examples and discuss them. I’m sure this happens somewhere, but I have not come across if so far.

What’s your take on this? Have I left out any important studies or perspectives? Do you have any other suggestions?
**Literature**


---

**An earlier version of this post appeared in Elisabet Tiselius' blog interpretings.**

**Recommended citation format:**