

Qualitative research: developing rigor

Prof Brett Smith discusses inter-rater reliability and developing rigor in qualitative research.

Qualitative research, once on the fringes in our field, now plays a central part in advancing knowledge in the sport and exercise sciences. Not only is qualitative research making up a large proportion of articles now published in many sport and exercise psychology journals. Grant funders and organisations like the World Health Organization are increasingly promoting qualitative methods as an invaluable way to co-produce research, produce contextually relevant knowledge and deliver impact. Despite such positivity, the rigor of this kind of research in our field requires urgent attention. This is because some of the most commonly used methods for ostensibly achieving qualitative rigor in our field are flawed and outdated, raising questions then about the quality of qualitative research produced and how we assess work.

Inter-rater reliability

Given space, I will first briefly focus on inter-rater reliability. That method is the most commonly used technique for claiming rigor in qualitative research within sport and exercise psychology journals. Inter-rater reliability, which is sometimes also termed investigator triangulation, first involves two or more researchers independently coding data. The same researchers then come together to compare codes and reconcile through discussion whatever coding discrepancies they may have. When a high level of agreement/consensus is demonstrated between the researchers the coding is deemed reliable, and thus the research displays rigor.

Despite the appeal, inter-rater reliability is ineffective as a method for delivering rigor. One reason for this is that humans, no matter how hard we try, cannot produce theory-free knowledge

Despite the appeal, inter-rater reliability is ineffective as a method for delivering rigor. One reason for this is that humans, no matter how hard we try, cannot produce theory-free knowledge. Because a researcher is incapable of stepping outside of his/her own history, culture, training and so on, the background knowledge held by the researcher inescapably influences coding. That is, we interpret data in ways that are dependent on us. For example, a cognitive exercise psychologist informed by self-determination theory would very likely interpret data when coding very differently from a discursive exercise psychologist who believes that selves are not innate but socially constructed through discourses and particular ways in which language is used during interaction. Any discussions over coding discrepancies between the two would also likely to be unresolved because of different theoretical positions they hold.

A further problem with inter-rater reliability relates to “who was involved” to establish it. It is very common for coders to be closely aligned professionally. For example, one coder may be a PhD student whilst the other is often the supervisor. We know however that power differentials, gender dynamics, nationality and past training can strongly influence inter-coder agreement here. For example, when discussing codes students might defer to their supervisors because they are believed to be the real experts. Another problem with inter-rater reliability is that it is always possible that coders might agree occasionally by chance. Also problematic is that there is no agreed upon threshold in the literature for what constitutes a numerically satisfactory level of agreement among coders to achieve reliability and more or less rigorous research. Is it 80%, 87% or 93% agreement? An examination of papers in sport and exercise psychology will reveal that what passes for an acceptable level of intercoder reliability

varies considerably according to the standards of different researchers as well as the method of calculation.

The numerous problems associated with inter-rater reliability (see Smith & McGannon, 2018) have led leading researchers to conclude that it is a myth (Morse, 1997), inappropriate for interpretive qualitative research (Levitt *et al.*, 2016), not worth pursuing (Braun & Clarke, 2013) and a flimsily retrofitted procedure in qualitative clothes to be avoided (Eakin, 2016). It is time then that sport and exercise researchers give up using inter-rater reliability as traditionally used for rigor purposes. Journal editors and reviewers should also consider this when judging qualitative research.

Developing rigor

How might then rigor be developed and the quality of qualitative research be judged? These questions are challenging not only because there rightly is no universally agreed criteria to develop or judge all qualitative research. Qualitative methods, methodologies and conversations about quality are expanding rapidly, thereby meaning researchers and reviewers need to keep abreast of developments and a vast literature. Expertise is not easily gained!

With that in mind, below is a list of criteria that *might* be used to enhance rigor or judge qualitative research. This list should not be thought of as fixed. Nor should it be applied in a predetermined and universal manner to all qualitative research.

The list of criteria can be added to and subtracted from as one goes about the practical task of developing rigor or judging each piece of qualitative research. Drawing on a range of articles on rigor (e.g. Morse, 2016; Smith, 2018; Tracy, 2010), the list of criteria one might draw on certain occasions to enhance rigor or judge it may include:

- *Worthiness* - is the topic worthy of study?
- *Substantive contribution* - does the work contribute to the field?
- *Epistemological and ontological alignment* - is the work epistemologically and ontologically coherent, from start to finish?
- *Prolonged engagement* - has the researcher spent quality and enough time collecting data?
- *Sampling* - has appropriate sampling strategies and sizes been used?
- *Informed methods choices* - too often interviewing or a thematic analysis are the default options to collect and analyse data in our field. Does the author demonstrate s/he made informed choices about the methods used?
- *Saturation* - has data or theoretical saturation been met? Is the researcher cognisant of the complexities behind claiming saturation?
- *Member reflections* - have participants provided feedback on data and interpretations? How did the researcher deal with that feedback?
- *Audit trail* - has there been transparency regarding ethics, data collection, analysis, etc?
- *Critical friends* - have other scholars provided critical feedback and encouraged reflection on ideas?

Within the sport and exercise sciences researchers need to be connoisseurs of the different ways to enhance rigor and judge research

- *Width and thick data* - is there evidence of appropriate and thick data?
- *Theory and rich interpretations* - does the study use sufficient, abundant, appropriate and complex ideas to create rich interpretations of data?
- *Generalisability* - qualitative research does not seek to be statistically generalisability. Nor should it be judged by that type of generalisation. But qualitative research may use other types of generalisability. Does the work, for example, display naturalistic generalisability, transferability or analytical generalisability?
- *Limitations* - are these correct and appropriate? Could the limitations have been resolved from the start?
- *Evocative* - is the paper evocative?

Conclusions

The standards for rigor are different to those used to judge the rigor and quality of quantitative research. To impose standards from quantitative research onto qualitative research (or vice versa) would then be an unintelligent move. But, qualitative research does still need to be held to high and very difficult standards. That means within the sport and exercise sciences researchers need to be connoisseurs of the different ways to enhance rigor and judge research. Whilst not easy, we need to remain up-to-date on the qualitative literature if we are to do excellent research and judge work well. The good news is that the UK has a growing number of researchers in our field with significant expertise to draw on. ■



Prof Brett Smith

Brett is the Head of Research in the School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham. He is former editor of *Qualitative Research in Sport, Exercise and Health* (Routledge) and co-editor of the *Handbook of Qualitative Research in Sport and Exercise*.

References:

- Braun, V. & Clarke, V. (2013).** Successful qualitative research: A practical guide for beginners. London: Sage.
- Eakin, J.M. (2016).** Educating critical qualitative health researchers in the land of the randomized controlled trial. *Qualitative Inquiry*, 22, 107-118.
- Levitt, H.M. et al. (2016).** Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology*. Advance online publication. doi:10.1037/qap0000082.
- Morse, J.M. (1997).** 'Perfectly healthy, but dead': The Myth of inter-rater reliability *Qualitative Health Research*, 7, 445-447.
- Morse, J.M. (2016).** Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, 25, 1212-1222.
- Smith, B. (2018).** Generalisability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health*, 10, 137-149.
- Smith, B. & McGannon, K. (2018).** Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*. Advance online publication. doi/10.1080/1750984X.2017.1317357.
- Tracy, S.J. (2010).** Qualitative quality: Eight 'big-tent' criteria for excellent qualitative research. *Qualitative Inquiry*, 16, 837-851.



Find the perfect candidate on the UK's most respected Sport and Exercise Science website

BASES website offers advertising for career opportunities to sport and exercise scientists across the UK and beyond. We're the first stop for sport and exercise related positions, internships, studentships, undergraduate courses and much more.

Our statistics show the BASES vacancies page has **over 13,000 visitors a month**.

Contact us to advertise your vacancy on our website and find your perfect candidate today.

For further information visit: www.bases.org.uk/Recruiters

All vacancies will be posted on our social media sites.

[@basesuk](https://twitter.com/basesuk) [@BASESUK](https://facebook.com/BASESUK) [bases_UK](https://instagram.com/bases_UK) [BASESUK](https://linkedin.com/company/BASESUK)

