Journal of Qualitative Research in Sports Studies

Volume 5, Issue 1, December 2011

Letting go of the side of the pool

\textsuperscript{a} John Metcalfe (University of Central Lancashire)

ISSN: 1754-2375
ISBN: 978-0-9566270-3-2
JQRSS Article No: 11/16-5-1-2011-PG[30]-057

To cite this article:

Self-archived URL link to this article:
https://www.academia.edu/3515222/JQRSS_Volume_5_Issue_1_December_2011

Advice to submitters - see JQRSS Guide to Contents & Open Call for Papers:
https://www.academia.edu/3513281/JQRSS_Overview_Guide_to_Contents_and_Editorials_by_Volume_-_Open_Call_for_Papers

JQRSS Mapping of Articles (2007- to date):
https://www.academia.edu/8827414/Mapping_of_JQRSS_articles_across_Undergraduate_and_Postgraduate_research

Copyright © Clive Palmer and the individual authors

Notice:
The discussions, statements of fact and opinions contained in the articles of The Journal of Qualitative Research in Sports Studies are those of the respective authors and cited contributors and are set out in good faith for the general guidance of student supported research and the promotion of pedagogical discussion in teaching and learning contexts. No liability can be accepted by the Editor, Editorial Board, Advisory Board, the reviewers or the authors/submitters for loss or expense incurred as a result of relying upon particular statements made or circumstances outlined in this journal.

Online Research Profiles:
academia.edu: https://uclan.academia.edu/ClivePalmer
ResearchGate: http://www.researchgate.net/profile/Clive_Palmer
British Conference of Undergraduate Research http://bcur.org/journals/
Letting go of the side of the pool

John Metcalfe
(University of Central Lancashire)

Abstract
This paper is a Post Doctoral reflective account about coming to terms with types of data that were valued, differently, during phases of a PhD study. The status of knowledge that a certain kind of data may provide began to be questioned and the tentative steps towards a quest for authenticity in data rather than statistical provability were being taken, shakily...

Naturally my understanding of research methods was based upon what I had experienced and been taught both as an undergraduate and as a postgraduate student. My interpretation of research integrity was primarily based on large sample sizes and statistical power. It was drummed into us that there was safety in numbers. Large numbers. Indeed, the larger the $n$-value the better, with the Holy Grail being clinical trials where sample sizes often stray into the hundreds of thousands. ‘How many participants have you got?’ and ‘what statistical tests are you doing?’ were typical questions us students would ask each other. Internal validity was also high on the agenda; controlling the independent variables so as to carefully monitor the dependent one. These teachings were neither wrong nor misguided per se, rather it was my interpretation, or the over emphasis I placed on them that caused my injudicious understanding. This folly was often reinforced when presenting at conferences or submitting to peer-reviewed journals ‘How many participants have you got?’ ‘What statistical analysis are you doing?’ and ‘how did you control for that variable? were once again asked.

This type of apprenticeship is not intrinsically bad; the methods can be very effective when used appropriately. However, I had become somewhat methodical in my methodology and I rarely strayed from this “big-numbers” approach. That is until my doctoral thesis. I had a problem. A big problem. In essence my research boiled down to monitoring elite mountain bikers during a 24 hour race. Given the nature of the sport, the population of elite mountain bikers is relatively small which in turn meant the sample size was inevitably going to be even smaller. Furthermore, for the data to be authentic it was imperative that they needed to be collected during a race. This presented me with a twofold issue. First, not many riders would be willing to comply with the testing protocol for fear that it may compromise their race performance, thus reducing the sample size even further. And secondly I would
not be able to control any variables due to the field testing nature of the research. This was new territory for me and was the cause of many a sleepless night.

The only previous study comparable to mine was by Laursen and co-workers (2003). These authors measured selected physiological variables from four riders for 12 hours during a similar mountain bike race. They set a precedent by using a repeated measures ANOVA to analyse the data. I was familiar with this approach and was the one I longed to do, however the authors did not make it clear how the data met the criteria for such a statistical test. The assumptions for ANOVA are that the data should be normally distributed, the variances for each condition are similar and data should be at least on an interval scale (Field, 2006). It is doubtful that four observations can be reliably tested for normal distribution parameters (Siegel and Castellan, 1988; Fallowfield et al., 2005) and thus I was sceptical of their use of ANOVA. Indeed many of my data sets had skewness and kurtosis ratios between -2 and +2 thus suggesting normality. However, this observation is more likely due to a statistical artefact rather than a normal distribution. Fallowfield et al. (2005) note that although ANOVA are relatively robust to violations, if such violations exist in every group (and they did with my data!) the reliability of the test is compromised, thus questioning the use and prophetic power of the test. Furthermore, when analysing my data sets it was apparent that for some variables the inter participant difference was comparable to the intra participant difference thus increasing the likelihood of a type II error. This was not looking favourable. More sleepless nights ensued.

In a relatively similar study Wirnitzer and Kornexl (2008) collected data on five males and two females during an eight stage mountain bike race. They correctly acknowledged the small sample size and employed nonparametric tests (Friedman test and Wilcoxon paired tests) to test differences. However, nonparametric tests discard information in the data by reducing them to ranks and it was the very nuances that make elite riders elite that I was looking for. Furthermore parametric tests are limited to one factor designs which did not suit the research design of my study. In addition I did not need to do any power tests to know that statistical analysis was not appropriate for the small number of subjects I would be studying – it was glaringly obvious it was underpowered (as most sports research is (Fallowfield et al., 2005)).

Clearly statistical analysis was not the way to go. However, like a novice swimmer not wanting to let to go of the side of the swimming pool I was reluctant to relinquish my perceived safety of statistical analysis. Furthermore some of the academics I sought advice from were wearing armbands and maintaining that statistical analysis was the appropriate strategy. I ploughed forward regardless.
The issue of internal validity had not gone away either. The paradox being that increasing internal validity would reduce the ecological validity and authenticity of the study. This dichotomy is often unavoidable, but the driving philosophy behind my research was for the outcomes to be applied and be accessible in a currency that is relevant to coaches and athletes. This was a key area that I did not want to compromise on. For me the data had to be authentic for it to be of any value. Fortunately this viewpoint is shared by other researchers. Hopkins *et al.* (1999) suggest that if a test of performance is to be of any use it should be analogous to the actual athletic event. Many laboratory protocols enable limiting physiological factors to be accurately measured, however they often have little relevance to sport (Krebs and Powers, 1989). Atkinson and Brunskill (2000) note that whilst simulated laboratory time-trials have been reported to be highly reliable for road cycling, they are valid only where there are no variations in external conditions (i.e. wind, hills, terrain, etc.) which, notwithstanding an enclosed velodrome, is an unrealistic scenario. Prins *et al.* (2007) reported a lack of relationship between a simulated laboratory mountain bike performance test and actual race performance. Furthermore, Hopkins *et al.* (1999) advise that a test should only be used if it is shown to be valid and more reliable than the actual event itself, and during which the participants are adhering to their normal training and nutritional practices. Failing these prerequisites they recommend that the “event itself provides the only dependable estimate of performance” (Hopkins *et al*., 1999: 472). Myburgh (2003: 182) advises that “when studying optimal training and performance it is probably better that researchers adapt their methods to suit the study of the relevant cohort”.

Because I was passionate about the authenticity of the data, freeing myself from controlling all but the dependent variable was an easier process than addressing the statistical elephant in the room. With regard to the latter I had concocted a hybrid design. I had settled on a process whereby I would perform a statistical analysis similar to Laursen *et al.* (2003) with the caveat that due to the low statistical power I would also interpret the data on a qualitative level. In short I was hedging my bets. This approach still did not feel right but it allowed me to sleep a little sounder for a few nights (though this may have been due to the consumption of red wine).

Of all things it was a chance meeting with Dr Beverly Hale, a highly regarded statistician, that finally convinced me to abandon the statistical analysis approach. She immediately pointed at the elephant and correctly informed me that statistical analysis is only a tool to aid our understanding of the data. And that in my case it was not the correct tool for job; a multiple case study design was. This design is commonplace in other branches of science where populations by definition are small (rare diseases, isolated psychological conditions, and animal populations at risk of extinction (Barlow and Hersen, 1984; Zhan and Ottenbacher, 2001)). Not employing
statistical analysis went “against the tide” of established practice in my field, but in my case was wholly justifiable. Intuitively it felt right and my normal sleeping pattern returned immediately (no need for red wine now).

There is no candyfloss moral to this story, though I did learn a valuable lesson. However it was not until several months after my viva voce that I realised I had learned it. As was often the case my focus had been on the endpoint – what empirical learning had happened? What content had I generated? Whilst this was an important aspect of the research I had also learned via the process that the process is just as important. I had wrangled and agonised for months over the correct analysis. In doing so I had unwittingly developed a sound justification for my choice of analysis and in doing so had galvanised my confidence in defending it. I had not only improved my swimming stroke, but I now had the confidence to use it and leave go of the side of the pool and successfully navigate the deep end.

References


**1st Reviewer’s comments:**

A qualitative convert? I think not. John's account illuminates some of the anxieties experienced when handling and attempting to analyse data whilst pondering how it will be received and used within the wider context of his research community. It appears that the defining point within his journey is not the end result that he initially anticipated but the methodological debate that ultimately enabled him to pick up the right tool for the job. For me, this is the Ph.D., it’s a learning process with more reflection and further questions than answers. I wonder if John will eventually dive into the pool.

**2nd Reviewer’s comments:**

Philosophically and methodologically it would appear that John and I are worlds apart. John’s early experiences of research revolved around ‘big numbers’, whereas mine revolved around using much smaller samples in order to, supposedly, gather ‘richer’ data. Despite our differential experiences, John and I have both experienced sleepless nights because of the apparently axiomatic ‘power of statistics’: John from trying to break from an approach that has shaped his thinking from the outset and underpinned much of the research he has conducted, and I from trying to gather and analyse so-called quantitative data for the first time in any meaningful way. Hopefully, like John, I too will reach the point when I am able to let go of the side of the pool and successfully navigate the deep end. If not, I find solace in the fact that red wine may help with the sleepless nights.