

# Book Review:

## *Quantitative Ethnography* by David Williamson Shaffer

Simon Buckingham Shum<sup>1</sup>

David Williamson Shaffer (2017). *Quantitative Ethnography*.  
Cathcart Press. ISBN-10: 0578191687. ISBN-13: 978-057819168

### Keywords

Quantitative ethnography, learning analytics methods, quantitative, qualitative, tutorials

<sup>1</sup>Email: [simon.buckinghamshum@uts.edu.au](mailto:simon.buckinghamshum@uts.edu.au) Address: Connected Intelligence Centre, University of Technology Sydney, Broadway, NSW 2007, Australia

The title is provocative: *Quantitative Ethnography*. Those two words are rarely uttered in the same breath except in conflict! For the former to be an adjective to the latter is surely ridiculous. Yet the “quant/qual” gulf is precisely what David Williamson Shaffer seeks to bridge. His message: people, we’re all on the same side, and I’m not asking you to sacrifice the principles you hold so dearly.

Sitting by the fireside, in conversation with someone you trust, is very much the feel of this book. It’s unlike any academic volume on research methodology that you’ll have read. In a conversational tone, packed with stories and humorous asides, this book will draw curious readers, from advanced undergraduates to professors. Shaffer accomplishes the extraordinary feat of introducing statistics and text analytics alongside ethnography and discourse analysis, gradually building the complexity, but without losing the reader. The way the ideas are presented has been honed from years of teaching complex material to graduate students, and it shows.

Shaffer’s one-line call-to-action is that he wants to move from Big Data to Big Meaning. In Chapter 1 he sets out his stall:

Fundamentally, this book is about how to use ethnographic techniques to guide statistical analyses of Big Data. At the same time, it explores how to use statistical techniques to increase the scope and power of ethnographic, and other qualitative methods of research.

It is a book about understanding why, in the digital age, the old distinctions between qualitative and quantitative research methods, between the sciences and humanities, and between numbers and understanding, limit the kinds of questions we can ask, in some cases, and lead us to accept superficial answers in others. *Quantitative Ethnography* is a research method that goes beyond those distinctions to help us understand how to make sense of our increasingly data-rich world. (p. 21)

Shaffer makes statements that every ethnographer, or other qualitative analyst would nod vigorously to. He is scathing about data mining that seeks to identify statistically significant patterns in an atheoretical way, and humorous with his examples (e.g., speed of moves is a statistically significant feature of chess grandmasters, so let’s tell beginners to move their pieces faster...).

We need a method for analyzing meaning to make sensible analyses of Big Data if we want to shed light on what people do and why. Put another way, the mountains of data that we see around us are more like a chain of islands in the middle of the cultural ocean of meaning. Ethnography is a way to chart that ocean, and *Quantitative Ethnography* is a way to use statistical tools to make better charts by finding landmarks amidst the mountains of data.

To do anything less — to pretend that the mountains of data are not themselves part of a sea of cultural significance — may be mathematically rigorous, but in the end is conceptually empty. (p. 19)

Shaffer argues for the importance of keeping in our sights not only the ability to describe the world through data, but to explain it; to understand why people do the things they do. In doing so, he highlights the risks of simply spotting patterns, without going further to make sense of those patterns to understand the world. Shaffer’s argument is both that ethnography can

inform how we make sense of these patterns, and that statistical approaches can “expand the tools of ethnography” to support, “the stories that we tell about the things people do, and reasons they do them” (p. 398).

How do we get there? It requires a new cadre of bilingual analysts (or teams of analysts who between them can do the translation work). On one side they’re fluent with the tools of statistics, data science and, increasingly, artificial intelligence, working with large amounts of “thin” data. On the other, they’re fluent also with the tools of thick data, coding layers of interpretation onto data of different modalities, from a deep understanding of how the culture being studied ascribes meaning to phenomena.

For “the quant readers,” Shaffer explains why we need to understand the activity giving rise to data traces, and how ethnography provides systematic methodological tools to code large amounts of data that are meaningful (Chapters 2–3). For “the qual readers,” there is an equally engaging journey into what different statistical concepts mean, illustrated with accessible examples (Chapter 4). Just as ethnographers handle the problem of bias in the ways they gather and make claims about phenomena, equally, statistics has developed an array of tools to handle this problem.

Now the bridge starts to be built (Chapters 5–7) — how to transform qualitative data into forms amenable to statistical analysis, and how to construct models. Chapters 8–9 go deeper, explaining where measures of inter-rater reliability fit into the story, and how to analyse the structure of meaning-making through a particular methodology and toolkit, Epistemic Network Analysis (ENA). A cunning device throughout is the use of the same kind of diagram to show how ethnography and statistics each seeks to make plausible claims about the world — a diagram that ultimately serves to show how they can be integrated (on page 253 for those seeking to jump to the punchline).

Chapter 10 is where Shaffer draws the threads together. The claim is that in the age of Big Data, “data is data” — there is no longer value in the prefixes qualitative or quantitative, when data from every modality can be treated computationally. We want analyses that are both rigorously grounded (with ethnography providing a disciplined way to ensure we do not decontextualize), and generalizable (with statistics providing the discipline for this). A key connecting concept is that of saturation, since both theoretical saturation in grounded analysis and sampling in statistics are warrants for making plausible, more general claims about what is “in the data.”

What does this work contribute specifically to the learning analytics (LA) community? As a truly multidisciplinary field, in which qualitative and quantitative colleagues are rubbing shoulders — and reviewing each other’s work — we need debate about the rigour of our methods for conducting large-scale yet grounded data analysis. Arguably, Quantitative Ethnography (QE) as an approach, with ENA as an exemplar tool, goes beyond multidisciplinary. It demonstrates what interdisciplinary advances can be made when two intellectual traditions are brought into deep dialogue. It’s not just a challenge for both sides of the divide to step out of their comfort zones and learn more about the other, but also provides a way forward with practical methodology and enabling software tools.

For readers specifically interested in discourse data and other event logs from collaborative learning, there is extra value since Shaffer works through examples of analyzing such data, in increasing depth, derived from his educational gaming/team simulation platforms. ENA will be of interest as an example of a practical QE methodology and software tool (Chapter 9: Connections) backed up with tutorials, code, a free hosted web app, and research case studies on <http://epistemicnetwork.org>. Shaffer is at pains to clarify that ENA is just an example of QE in action; the book is designed to support the community in using QE across other contexts, with Shaffer making the claim that

[M]any of the principles and ideas that are the foundation of Quantitative Ethnography will apply equally well to other approaches for making sensible interpretations of Big Data.

For one thing, when we combine ethnographic and statistical tools to analyze Big Data, we also get a larger set of tools for making sense of smaller data: the kind of data that is used by ethnographers, and historians, and journalists, and a host of scholars who study all manner of art, literature, and social interaction. Making meaning of Big Data gives us insight into how to use statistics to understand cultural material of all kinds, and the techniques of Quantitative Ethnography work quite well for Quantitative History, or Quantitative Journalism, or Quantitative Literary Analysis. (p. 20)

The learning analytics community is of course interested not only in rigorous methodologies for researchers to model and analyse rich forms of learning, but also in practical tools that can be used by educators and learners. The book focuses on the former, but in the concluding chapter, we glimpse an infrastructure of the sort that will immediately resonate for LA readers. Shaffer introduces a real-time analytics dashboard for educators, one that automates ENA on student chats and suggests interventions based on the model. As such, this represents an advanced, real time, discourse-centric learning analytics system, with the most recent work presented at LAK ’18 (Herder et al. 2018). However, this dashboard has yet to be deployed and evaluated at scale, something that this readership will be keen to see.

What does the book omit? In brief, detailed instructions on how to conduct statistics, and how to conduct ethnography. Those might seem like important oversights, but as Shaffer states in the introduction, one book cannot cover everything:

[T]his is not a book about ethnography, any more than it is a book about statistics. We will look at some important statistical issues, and also discuss some of the important principles of ethnography — and of

interpretive or qualitative research more generally. But fundamentally this is a book about how to use ethnographic techniques to guide statistical analyses of Big Data. And at the same time it is an exploration of how to use statistical techniques to increase the scope and power of ethnographic and other qualitative methods of research. (p. 20)

Throughout, and particularly in each chapter's concluding "next steps" section, Shaffer refers the reader to both primers and more advanced reading. The book's goal is to help translate across the great divide, so that quantitative and ethnographic readers can better understand what the other does, and importantly, why. The reader will find the introduction to statistics very engaging, but must then hone their skills in using its tools to perform such analyses using other training resources. Similarly, this book does not explain how to conduct an ethnographic field study, in education or any other context.

QE as an approach is, as far as I can tell, agnostic about the particular school of ethnography to which one subscribes (for a brief history of ethnography, in the context of how it contributes to the design and analysis of human-centred technology, see Dourish, 2014). Ethnographers would be rightly outraged at the idea that the heart of what they do can be quantified. But this is not Shaffer's claim (perhaps the book's title is too provocative in this regard). For ethnographic and other qualitative researchers, the book's story starts when, via whatever methodology you use, you are ready to begin working with your data: the book now invites you to consider adding to your toolkit the analytical tools of statistics. I do, however, anticipate that qualitative researchers will have more trouble than their quantitative colleagues in accepting the book's arguments. I suspect that some will still see a positivist wolf in sheep's clothing, but that is an argument for them to mount. Reviews of the book from the ethnography community would indeed be very interesting to read.

Curiously, then, the book is suited for three very different readerships. First, those who consider themselves newcomers to either or both statistics and qualitative approaches, since it really assumes no prior knowledge. Second, those who are already fluent in one worldview, and want to better understand the other. Third, those who are already deeply interdisciplinary, and are curious to compare and contrast their resolutions of the tensions with Shaffer's approach. Perhaps it will leave the latter two groups dissatisfied that he has not engaged their more nuanced concerns in greater depth, but it seems reasonable for the book to leave that dialogue for future venues and volumes.

To conclude, QE is an approach to analyzing the huge amounts of data now available to us about socially rich contexts, but doing so in a way that seeks to honour the rigour and discipline developed over the years in qualitative research, ethnography in particular. Academically, the learning analytics community should consider this one of its reference points. Moreover, given its engaging, accessible style, it is highly recommended for readers in advanced undergraduate, masters, and doctoral programs as a way to illuminate the quantitative and qualitative worldviews that must be brought together in a human-centred, data-intensive discipline.

NOTE: To see Shaffer present the key ideas of the book in 45 minutes, in the same tone in which he writes, see his LAK '18 keynote address: <http://bit.ly/ShafferKeynoteLAK18> and the extended Q&A session that followed: <http://bit.ly/ShafferDiscussionLAK18>

## Declaration of Conflicting Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author declared no financial support for the research, authorship, and/or publication of this article.

## References

- Dourish, P. (2014). Reading and interpreting ethnography. In J. Olson & W. Kellogg (Eds.), *Ways of Knowing in HCI* (pp. 1–23). Springer Verlag. [https://doi.org/10.1007/978-1-4939-0378-8\\_1](https://doi.org/10.1007/978-1-4939-0378-8_1)
- Herder, T., Swiecki, Z., Fougat, S. S., Tamborg, A. L., Allsopp, B. B., Shaffer, D. W., & Misfeldt, M. (2018). Supporting teachers' intervention in students' virtual collaboration using a network based model. *Proceedings of the 8<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK '18)*, 5–9 March 2018, Sydney, NSW, Australia (pp. 21–25). New York: ACM. <http://doi.org/10.1145/3170358.3170394>