Defending interpretivist knowledge claims in engineering education research

Scott Daniel, Llewellyn Mann, Alex Mazzolini
Swinburne University of Technology
sdaniel@swin.edu.au

SESSION C5: Systems perspectives on engineering education

CONTEXT In interdisciplinary research, tacit epistemological differences can influence how research is interpreted and judged as trustworthy or otherwise. One example is in education research in engineering. A complication in the development of engineering education research as a field is that many of its practitioners have moved into education research from a background in traditional engineering, underpinned by a positivist epistemology with established criteria of research rigour. However, an arguably similar consensus has not been reached for criteria of research quality in education, at least not in inter-disciplinary areas like engineering education. One consequence is that researchers from such a positivist tradition can be dismissive of interpretivist research findings, and only find positivist research trustworthy.

PURPOSE How to defend interpretivist knowledge claims in engineering education research?

APPROACH Walther, Sochacka, and Kellam (2013) used an analogy with quality management in engineering to develop a process-oriented framework for interpretive research quality. Instead of judging only the quality of research outcomes, as is typical in positivist research, they focused instead on the processes of both making and handling data.

In this paper, this framework is unpacked and used to defend the results of the authors' previously published phenomenographic study of lecturing (Daniel, 2016; Daniel, Mann, & Mazzolini, 2016).

RESULTS In this paper, the reliability and validity of the outcomes of a previous phenomenographic study of ways of experiencing lecturing are established. This is achieved through reference to established conventions in phenomenographic research, thick descriptions of how the data was collected and analysed, and comparison to the results of similar studies, all within the framework of interpretivist research quality developed by Walther et al. (2013). Such thick descriptions of data collection and analysis are often omitted from phenomenographic publications, whereas detailing this process can lend weight to such research's reliability.

CONCLUSIONS Interpretivist methodologies have an important role in engineering education research. By taking pains to establish the validity and reliability of interpretivist research outcomes, it is hoped they will be accepted more widely amongst researchers, regardless of whether they come from a positivist or interpretivist background.

KEYWORDS Research quality, epistemology, interpretivist

Introduction

"Show me a cultural relativist at 30,000 feet and I'll show you a hypocrite"
Richard Dawkins
River out of Eden (1995)

In inter-disciplinary research, tacit epistemological differences can influence how we interpret research and judge its trustworthiness. One example is in engineering education research, and in STEM education research in general. A complication in the development of STEM education research as a field is that many of its practitioners have moved into education research from a background in traditional science, underpinned by a positivist epistemology with established criteria of research rigour. However, a similar consensus has not been reached for criteria of research quality in STEM education research. One consequence is that researchers from a positivist tradition can be dismissive of interpretivist research findings, and only find positivist research trustworthy. This is illustrated in the above quote, and in the following excerpt from an interview conducted with a physics lecturer (Daniel, 2016):

> *I went to a talk by Eric Mazur that made me more aware that there is actually not just some theories on why active learning might be better, but a lot of hard-nosed detailed statistically significant research, at first-year level anyway, on why it is better … and that was really what made me aware that* ***this isn't just teaching and learning specialists wittering on about the latest pedagogical craze, this is well backed by hard evidence with good p-values*** *[Zorro, p. 32]*

In a positivist tradition, research quality is typically judged by the validity and reliability of findings. Validity can be defined as the "agreement of the results of a measurement with the true value of the measured quantity" and reliability as the "repeatability" of measurement (Sirohi & Radha Krishnan, 1983). With their emphasis on measurement, validity and reliability are sometimes operationalized as accuracy and precision.

But how to make sense of these concepts in interpretivist research, where there are no objective 'true values', and the complexity and uniqueness of social systems belie the possibility of exact repeatability?

In social science research, there is a long tradition of exploring these questions of interpretivist research quality (e.g. Guba (1981), Krefting (1991), and Schwandt, Lincoln, and Guba (2007)). However, in this study we used a new framework for research quality that is perhaps more appropriate and accessible for engineering educators, as it attempts to build a bridge between engineering practice and interpretivist research.

Walther et al. (2013) used an analogy with quality management in engineering to develop a process-oriented framework for interpretive research quality. Instead of judging only the quality of research outcomes, as is typical in positivist research, they focused instead on the processes of both making and handling data. They reframed reliability as the extent to which random influences on the research process are minimised, and unpacked validation into four different aspects, centred around the question of "whether the researcher sees what they think they see" and how they conform to meaning conventions in reporting their work to the relevant research community.

In this paper, we will describe this research quality framework and give contrasting examples of how it can be used to characterise quality interpretivist research. Then we will explore how it was used to defend the first author's PhD phenomenographic research into lecturers' different ways of experiencing lecturing, in the epistemological cold-war battleground of engineering education.

Reliability and Validity

Scientific Research Quality

Scientific research quality is generally evaluated by its reliability and validity, operationalised as precision and accuracy (Figure 1). The goal is to have results clustered tightly around the centre of the metaphorical target. That is, results that are both precise and accurate, as represented in the bottom right diagram.

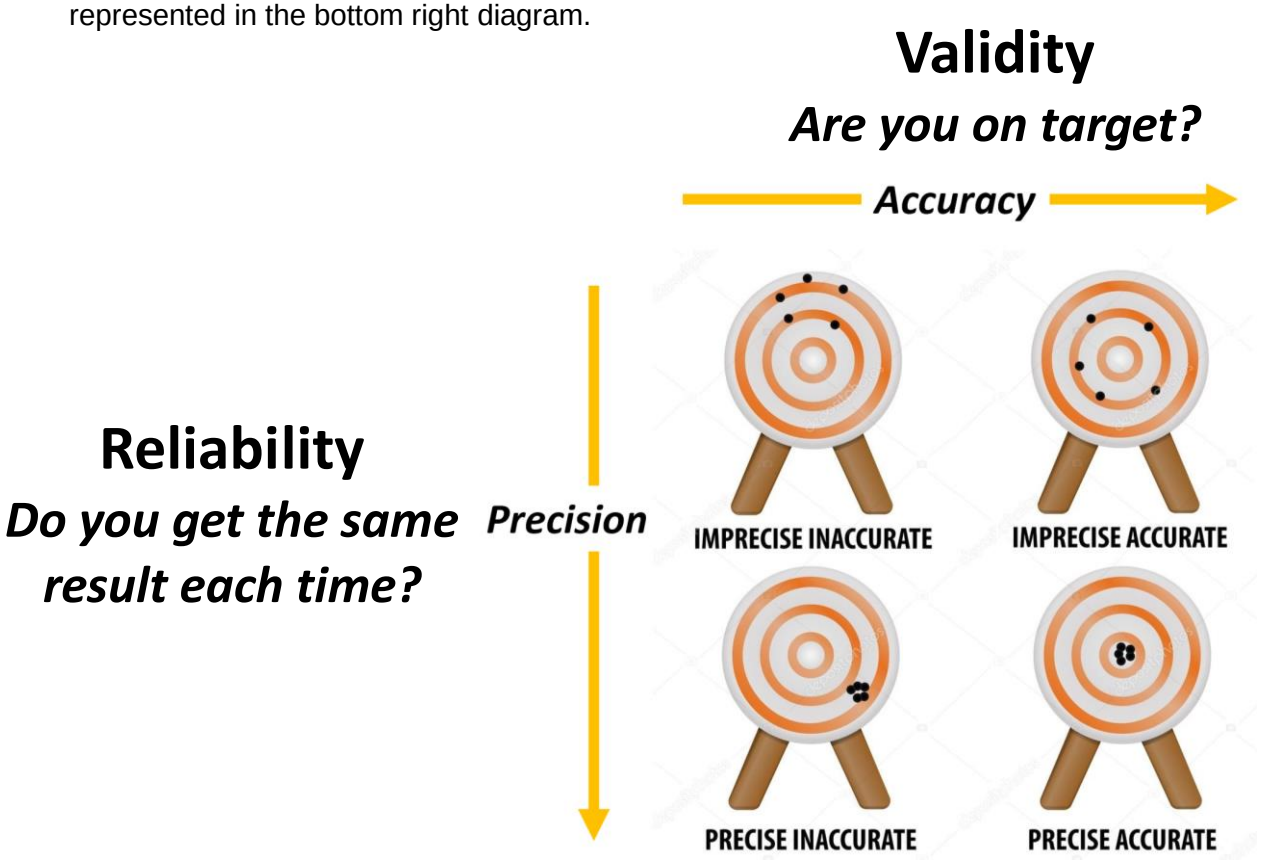

Figure 1: Characterising scientific research quality

Results that are on-target but only clustered loosely, as in the top right, are said to have a large random error. This could reflect a low-resolution measuring instrument, or perhaps a relationship affected by factors you haven't considered. For example, although a person's height is a useful predictor of their weight, the impact of other relevant variables means that if you measured the weight of a number of individuals with the same height, there would still be considerable variation.

Conversely, results that are tightly clustered but off-centre, as in the bottom left, indicate either a zero or systematic error. One recent high-profile example of such precise but inaccurate measurements was the six-sigma result of neutrinos traveling faster than light (Adam et al., 2012), which was later found to be spurious due to a subtle systematic error. Another example is the crash of the NASA Mars Climate Orbiter in 1999, because one team was measuring thrust in imperial units, but another team was assuming these values were in metric units (Grossman, 2010).

Research questions like "What is the speed of neutrinos in a vacuum?", or "What is the relationship between height and weight of Australian adults?", are about investigating objective reality and collecting 'hard' data. However, typical education research questions, like "What does 'great teaching' mean to different people?" or "How do different students perceive successful research supervision?" are instead about the researcher making sense of subjective experiences. That is, they reflect an interpretivist, rather than positivist,

epistemology. If the same criteria of positivist research quality are applied to interpretivist education research, asking questions like “Do you get the same result each time?” and “Are you on target?”, one may be tempted to conclude, like Zorro in the above quote, that such research is meaningless and “just teaching and learning specialists wittering on about the latest pedagogical craze”.

Characterising interpretivist research quality

Walther et al. (2013) developed a framework for interpretivist research quality, adapting the concepts of positivist research quality to the interpretivist domain (Table 1). Instead of the positivist focus on judging only the quality of research results, they applied an analogy with quality management and focused on the research processes of making and handling data.

Table 1: Frameworks of research quality

	Positivist research	***Interpretivist research*** ***(Walther, Sochacka et al., 2013)***
Focus	Results	*Processes* *cf. Quality management*
Reliability	Do you get the same result each time? **No random error**	*Mitigating random influences on the research process* ***Process reliability***
Validity	Are you on target? **No systematic error**	*Does the researcher see what they think they see?* ***Theoretical, procedural, communicative, and pragmatic validation***

They interpreted reliability as ‘process reliability’, and unpacked validation into four different aspects (Table 2). Note that they used the term ‘theory’ to mean the researcher’s interpretation or ‘sense-making’ of the phenomenon under investigation.

Table 2: Four different aspects of validation (Walther et al., 2013)

Aspect	**Related to:**
Theoretical	the fit between the social reality under investigation and the theory generated
Procedural	features of the research design that inherently improve the fit between the reality studied and the theory generated
Communicative	the integrity of the interlocking processes of social construction with the relevant communication communities
Pragmatic	the compatibility of theoretical constructs with empirical reality

As in positivist research, the goal is always to conduct valid and reliable research. To help demonstrate what valid and reliable interpretivist research looks like, we will first show what it is not (cf. variation theory (Bussey, Orgill, & Crippen, 2013)), by giving some counter-examples of low-reliability and low-validity research.

Low reliability example

If, for example, a researcher had to go through a dozen research assistants before finding one that agreed with her thematic coding of some interview data (as has been reported anecdotally), the research would have low process reliability. The analysis would arguably be more a reflection of the idiosyncrasies of the researcher, than the views put forward in the interviews.

Low validity example

If a researcher were to use student feedback surveys with the belief that this was a measure of teaching quality, this would be an example of low validity research – the researcher would not be seeing what they thought they were. Student feedback surveys are more a reflection of the presenter's charisma or fluency (Carpenter, Wilford, Kornell, & Mullaney, 2013; Naftulin, Ware, & Donnelly, 1973), or the respondents' biases (MacNell, Driscoll, & Hunt, 2014), than teaching quality or student learning.

Low reliability and low validity example

Were a researcher to investigate what it means for research supervision to be a success, by interviewing students, including some of her own, about their perceptions (as has been reported anecdotally), this would be an example of research that is both low reliability and low validity. Because of the unexamined power dynamic between the supervisor and her students, the research would not be valid. Further, because of the mix of students, with some being her own, and some not, this would arguably be an 'unmitigated random influence on the research process', or in other words meaning that the research had low reliability.

As opposed to these low reliability and low validity examples, in the following section we argue for the high reliability and high validity of one of our previous studies.

Establishing the reliability and validity of our previous work

We previously conducted a phenomenographic investigation of experiences of lecturing, asking 'What are the different ways of experiencing lecturing?'. Although other studies have investigated different experiences of teaching in general, this was the first study with a specific focus on lecturing. The results of that investigation are explored in detail elsewhere (Daniel, 2016; Daniel et al., 2016). In summary, we identified the following five qualitatively distinct ways of experiencing lecturing, framed by three themes of experiencing awareness: student diversity, interaction, and lecture purpose.

1. Lecturing as soliloquy
2. Lecturing as connecting meaning
3. Lecturing as cultivating individuals
4. Lecturing as transformatively co-creating
5. Lecturing as enacting research

Claims of research quality in engineering education often remain tacit. In the first author's PhD thesis (Daniel, 2016), the five criteria of the Walther et al. (2013) quality research framework were explicitly addressed in multiple ways, for both *making* and *handling data*.

In the following sub-sections, representative examples of how this was achieved are given for each of the criteria, to give a flavour of how this framework can be used in practice. For the sake of brevity, in each case the quality of the research process of only either *making data* or *handling data* are discussed. The first person 'I' is used to indicate it was the first author's analysis and interpretation.

Theoretical validation

Walther et al. (2013) describe this quality criterion (p. 640) as answering:

> *Do the concepts and relationships of the theory appropriately correspond to the social reality under investigation?*

In other words, this aspect focuses on the question: to what extent does the knowledge produced by an investigation relate to the empirical phenomenon in question?

Making the data

In our study, this was addressed by purposely recruiting as diverse a sample as possible. In phenomenography, the goal is not to gain a representative sample, but instead to capture a wide pool of experiences of the phenomenon in question. For this investigation into lecturing, the dimensions along which we sought to maximise diversity included gender, university context (regional versus urban, research-focused versus technology-focused), discipline, and years of lecturing experience.

Furthermore, the object of study was not lecturing practice *per se*, for which perhaps an observational study would be most appropriate. Instead of such a first-order empirical study, the object of study was understanding the different ways in which lecturing is *experienced*. Although the extent to which it can do so is innately limited (Säljö, 1997), the best tool we have for this is analysing the different ways people talk about the phenomenon, to deconstruct what is salient to them about the phenomenon. Thus, semi-structured interviews were used.

Procedural validation

Procedural validation is about making clear what aspects of the research design improve the fit between the social reality and the interpretation thereof.

Handling the data

In analysing the transcripts it was important to try to identify instances in each transcript, and in the pool of transcripts as well, of each critical variant to ensure that my interpretation was not based on some idiosyncratic analysis of one decontextualised utterance but grounded in the context of the transcript and the pool of transcripts.

In addition, when I had felt I had identified some ‘essence’ of a transcript after reading it, I made sure I could identify supporting quotes to defend my knowledge claim. I had to always question my interpretations, and purposely look for and consider disconfirming instances, to limit the extent to which I was projecting my biases on to the data. This process of cyclically evaluating conjectured interpretations against the data is sometimes called the ‘constant comparative method’ (Glaser & Strauss, 1967).

I used a number of strategies to critique my own interpretation and decision-making process of analysis. I kept a detailed record (cf. Chapter 6 of my thesis) of decisions and interpretations and sought always to evaluate them against the transcripts and disregard intuitive interpretations that I could not defend without quotes. I was aided in this process by two critical friends (Costa & Kallick, 1993) who helped shape the analysis by challenging me on points that I had glossed over or not backed up with supporting quotes. If I could not argue from the quotes for a particular interpretation, it was disregarded.

Communicative validation

This criterion refers to the extent to which meaning and interpretation is communicated with different stakeholders in the research process: participants, the research team, education practitioners, and other researchers and the literature more broadly.

Making the data

At the beginning of each interview, through both the consent form and a quick spoken introduction, I would describe my research project and highlight that I was interested in *their* experiences of lecturing, and that there were no ‘correct’ answers. If participants asked what I meant by lecturing, I would explain that I was asking about what happened when they had a

lecture on their timetable, or about what happened in the lecture hall. In this way, I ensured that we were talking about the same phenomenon, but avoided projecting any of my own judgements or understandings about lecturing.

As I wrote on my interview protocol as a prompt for myself, I would 'guard against assuming any terms they say'. In practice this meant that I would avoid paraphrasing participants' ideas back to them to seek their confirmation, which would involve recasting their ideas through my awareness, or not checking terms at all, which would imply me making assumptions about meanings of terminology. Instead, when relevant concepts were referred to, I would neutrally probe them, using questions such as "what do you mean by that?" or "can you give me an example?", with the assumption in the interpretation that the provided example is an exemplar of that concept, that for the respondent it epitomises the features of the concept important to them. Sometimes, I would simply repeat their words with an upward inflection – a non-judgemental way of asking them to elaborate on the meaning of a particular term or phrase.

Using these strategies I communicated with participants my motives for the research, and clarified any ambiguous terminology.

Pragmatic validation

Walther, Sochacka et al. (2013) characterise pragmatic validation as the "process of determining whether the theory and constructs used or developed in a particular study can withstand prolonged exposure to the empirical reality" (p. 647). That is, do the results actually make sense. In phenomenography for example the goal is to describe variation, therefore the results should actually show some variation.

Handling the data

My analysis was pragmatically validated in several ways.

The analysis process was meaningful for me
Similar to how the participants found the interview process a useful reflective device, as a lecturer myself, I found the interviews and subsequent analysis a prompt for reflection on my own practice and understanding. It also prompted me to reflect on issues outside of teaching, as well as helping me make sense of other education contexts in new ways. I explore these reflections in detail in my thesis, but overall I can assert that the process has been meaningful for me.

Phenomenographic assumptions validated by findings
Phenomenography assumes that there is a coherent hierarchy of categories of description that relates the variation in how participants' transcripts reflect the different ways they experience a particular phenomenon. I found such a hierarchy, and therefore my study is pragmatically validated.

Potential application to professional development
Beliefs about teaching are a necessary, but not sufficient, component of successful pedagogical reform (Henderson, Beach, & Finkelstein, 2011). It is my hope that this study will contribute to the discussion about teaching beliefs in a meaningful way by prompting lecturers to reflect on their practice, and perhaps eventually be incorporated into future effective professional development programs for lecturers, thereby incidentally demonstrating its pragmatic validity.

Process reliability

Within a positivist epistemology, reliability refers to the consistency of repeated measurements. In an interpretivist paradigm, the complexity and uniqueness of different participants and contexts are central, and so the criterion of repeatability is no longer

applicable. Instead, Walther et al. (2013) adapt the idea of reliability into interpretivist research as the extent to which the research process is independent of random influences, including the idiosyncrasies of the researcher. They advocate for the "development and explicit documentation of dependable procedures in making and handling the data" (p. 649).

Making the data

I provided 'explicit documentation' of my phenomenographic data collection process in my thesis (Daniel, 2016), and summarise it briefly here.

I conducted two pilot interviews, which were not used in the analysis, which I recorded and reviewed with my supervisors to refine the interview protocol and my interview technique. When debriefing with one of the pilot interview participants, and analysing with him to what extent I had allowed my own awareness to influence the direction it took, he commented poetically that "you opened a canvas for me to paint my understanding on".

I recorded the interviews on a digital voice recorder, then had them transcribed by a professional transcriber (except for two interviews which I transcribed myself), and then subsequently verified the transcription myself, to correct phonetic substitutions or other transcription errors.

Conclusion

Vouching for the quality of interpretivist research processes is sometimes overlooked compared to the review processes in place for judging research outcomes, typically published as conference or journal papers. Quality research outcomes are predicated upon quality research processes, but claims of the latter are most often implied rather than made explicit in engineering education research. This work makes an important first step in interpretivist engineering education research by using a systematic quality framework, developed through an analogy with engineering quality management, to explicitly argue for the reliability and validity of a phenomenographic education research study.

References

Adam, T., Agafonova, N., Aleksandrov, A., Altinok, O., Sanchez, P. A., Anokhina, A., . . . Autiero, D. (2012). Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. *Journal of High Energy Physics, 2012*(10), 1-37.

Bussey, T. J., Orgill, M., & Crippen, K. J. (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. *Chemistry Education Research and Practice, 14*(1), 9-22. doi:10.1039/C2RP20145C

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychon Bull Rev*. doi:10.3758/s13423-013-0442-z

Costa, A. L., & Kallick, B. (1993). Through the lens of a critical friend. *Educational Leadership, 51*, 49-49.

Daniel, S. A. (2016). *Experiences of lecturing.* (PhD), Swinburne University of Technology, Melbourne. Retrieved from http://hdl.handle.net/1959.3/422498

Daniel, S. A., Mann, L. M. W., & Mazzolini, A. P. (2016). *A phenomenography of lecturing*. Paper presented at the 44th SEFI Conference, Tampere, Finland. http://sefibenvwh.cluster023.hosting.ovh.net/wp-content/uploads/2017/09/daniel-a-phenomenography-of-lecturing-56_a.pdf

Dawkins, R. (1995). River out of Eden: A Darwinian View of Life, Science Masters Series: London: Weidenfeld & Nicholson.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*: Transaction Publishers.

Grossman, L. (2010). Metric math mistake muffed Mars meteorology mission. *Wired*.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development, 29*(2), 75-91.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching, 48*, 952-984. doi:10.1002/tea.20439

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American journal of occupational therapy, 45*(3), 214-222.

MacNell, L., Driscoll, A., & Hunt, A. (2014). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, 1-13. doi:10.1007/s10755-014-9313-4

Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). Doctor Fox Lecture - Paradigm of Educational Seduction. *Journal of Medical Education, 48*(7), 630-635.

Säljö, R. (1997). Talk as Data and Practice — a critical look at phenomenographic inquiry and the appeal to experience. *Higher Education Research & Development, 16*(2), 173-190. doi:10.1080/0729436970160205

Schwandt, T. A., Lincoln, Y. S., & Guba, E. G. (2007). Judging interpretations: But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation, 2007*(114), 11-25.

Sirohi, R. S., & Radha Krishnan, H. C. (1983). *Mechanical Measurements* New York, NY: Wiley.

Walther, J., Sochacka, N. W., & Kellam, N. N. (2013). Quality in Interpretive Engineering Education Research: Reflections on an Example Study. *Journal of Engineering Education, 102*(4), 626-659. doi:10.1002/jee.20029