International Journal of Engineering Education Vol. 32, No. 6, pp. 2378–2394, 2016
Printed in Great Britain

0949-149X/91 $3.00+0.00

Standards for Quality of Research in Engineering Education*

JONTE BERNHARD
Linköping University, Norrköping, Sweden. E-mail: jonte.bernhard@liu.se

CAROLINE BAILLIE
University of Western Australia, Perth, Western Australia, Australia. E-mail: caroline.baillie@uwa.edu.au

The understanding of quality in scientific work is fundamental and determines what researchers judge to represent reliable knowledge in their field. Although quality criteria are used daily in research, there are few extensive discussions available especially in Engineering Education Research (EER). For the development of future high-quality research in EER we argue that it is necessary that the EER-community begin to negotiate criteria for quality. In this paper we propose tentative criteria with a special focus on qualitative EER, although we argue that several of our proposed criteria are also appropriate for quantitative EER.

Our proposed criteria are divided into three main categories: quality of a study in general, quality of the results and validity of the results. We describe these in detail, together with a number of subcategories for each and introduce a hypothetical study to exemplify our criteria. It is stressed that the proposed criteria are tentative and that criteria need to be open for debate and need to evolve as research evolves.

Keywords: engineering education research; research quality; quality criteria

1. Introduction

The understanding of quality in scientific work is fundamental, and determines what researchers consider reliable knowledge in their field. In recent years, studies into critical factors for learning in engineering education, i.e. Engineering Education Research (EER), have emerged worldwide [1, 2]. However, although quality criteria are used daily in research, extensive holistic reviews and position papers discussing research quality in EER have been lacking. Without a thorough discussion related to all aspects of research quality we risk the use of random and subjective criteria. This can lead to an unquestioning acceptance of one research paradigm, or epistemology, which is often the dominant paradigm. Furthermore, we see an inappropriate matching of research questions to methodologies, confusing reliability with validity, and a lack of transparency in the criteria being applied to what constitutes quality in research. For these reasons we felt it was important to initiate a discussion within the EER-community about quality criteria. First versions of the criteria discussed in this paper were presented at the Nordic Network for Engineering Education Research (NNEER) seminar in Aalborg, Denmark, in 2012 and at the SEFI annual conference in Thessanoliki [3], Greece, in 2012. A revised version [4] was presented at the Research in Engineering Education Symposium (REES) in Kuala Lumpur, Malaysia, in 2013. This paper presents a further revision of the criteria presented previously, amended on the basis of feedback and with more examples. From the need we have seen for a deeper discussion and raised awareness regarding quality issues, we note with satisfaction that, in parallel to our work, Walther et al. [5] have developed quality criteria for interpretive EER, as discussed below.

The paper is organised as follows: In section 2 we discuss some critical features within the debate relating to criteria for quality, in section 3 we present a brief review of quality criteria discussed in the literature and in section 4 we present our proposed criteria. We also introduce a hypothetical study that is used to exemplify our criteria.

This paper aims to contribute to a continuous and (hopefully) ongoing debate about quality criteria in EER. It is beyond the scope of this paper to present a complete review of all work related to research quality published in education research literature. Although what we communicate in this paper will be relevant to quantitative research, we focus on qualitative research since most controversies related to research quality in EER have been, rightly or wrongly, focussed in this area.

2. Critical features in the debate about criteria for quality

2.1 Methodology and epistemology

Epistemology is a branch of philosophy concerned with the theory of knowledge. It attempts to provide answers to the question, 'How, and what, can we know?' This involves thinking about the nature of

* Accepted 18 June 2016.

knowledge itself, about its scope and about the validity and reliability of claims to knowledge. Research methods provide ways of approaching, and hopefully answering, our research questions. However, first we need to identify our goal and be able to justify our choice. We need to be clear about the objectives of our research and we need to have a sense of what kinds of things it is possible for us to find out. In other words, we need to adopt an epistemological position. [6, p. 2]

In order to study learning, we need to think about what we mean by the terms 'learning' and 'knowledge'. This will influence the underlying theory we choose to adopt and the way we conduct our research. Epistemology is the term used to describe *ways of knowing*. Recently it was pointed out:

> Some . . . interpretations of the Gold Standard have *privileged a single research approach and type of evidence* regardless of the development of the problem space, specific research question, available technologies and instrumentation, and cost or ethical considerations. If such *interpretations of this policy exclusively privilege [randomized controlled trials] and quantitative evidence, it would disregard high-quality, qualitative research approaches and other contemporary approaches* and, thus, the evidence flowing from such inquiries. [7, p. 5, our italics]

The above quote suggests to us that high quality in research requires that the approach, and type of evidence collected, be aligned to the "the problem space, specific research question, available technologies and instrumentation, and cost or ethical considerations". However, Case and Light [8] argue that "methodological decisions need to be more explicitly represented in reports . . . in engineering education research", and in their review and meta-analysis, Koro-Ljungberg and Douglas [9] note that the issue of methodology has received limited explicit discussion in EER-literature and that many studies lack epistemological consistency.

Recently, Baillie and Douglas [10] have compared studies related to engineering design from five different groups of authors in terms of their respective epistemologies, methodologies, and methods. These different approaches enable different kinds of results and what can be seen. Baillie and Douglas take the view that none of these approaches are "right" or "wrong"; on the contrary, they argue that all studies contribute to a full understanding of the design process and its impact.

We agree with Baillie and Douglas [10] that all epistemological and methodological positions are needed. Hence, we maintain that methodological and epistemological awareness and consistency are clearly examples of important factors contributing to the quality of an EER research paper but which appear to be rarely discerned.

2.2 Method— versus problem-led research

Streveler and Smith [11] define "rigorous research in engineering education by using the guidelines provided by the [US] National Research Council (NRC) in *Scientific Research in Education*" [12]. According to the NRC, scientific research in education should:

1. Pose significant questions that can be investigated empirically
2. Link research to relevant theory
3. Use methods that permit direct investigation of the question
4. Provide a coherent and explicit chain of reasoning
5. Replicate and generalize across studies
6. Disclose research to encourage professional scrutiny and critique

Nevertheless, Borrego and Bernhard [2] note that the NRC's report [12] is situated within the "curriculum studies" (as described below) tradition and furthermore, these criteria are not quality criteria, but criteria describing what would count as *scientific* research in education. They have also explored differences in US-based "curriculum studies" and EU "didaktik" traditions [13, 14]. In the "didaktik" tradition *what* should be learned and *why* it should be learned is an important question, and quality of instruction is judged in terms of the professional appropriateness of knowledge and skills learning. "Bildung", i.e. the formation of the individual as a whole beyond knowledge and skills, is also deemed an important concept [15]. In the "curriculum studies" tradition, *how* a given topic is best taught is an important question, with the quality of instruction judged in terms of student achievement on (objective) tests and student perception of classroom activities. Borrego and Bernhard relate these differences to Bishop's [16] analysis of mathematics education research, in which he explains that differences in conceptions of quality in research can be traced back to "problem-led" and "method-led" educational research traditions. According to Bishop, problem-led researchers "would judge research quality by what light is shed by the research on the problem under consideration," in other words, "the quality of the ideas and insights generated" (p. 716). In contrast, a "method-researcher would argue that, without methodological criteria, such research is worthless" (p. 716). Method-led researchers would judge research quality in terms of the proper use of methodology and quality of the evidence.

For research to be of high quality, we maintain that it needs to be problem- (cf. section 4.2 and section 4.3) as well as method-led (cf. section 4.2.

and section 4.4). However, as pointed out by Gadamer [17, pp. 226–227], the rigorous application of a method does not guarantee productive and good research, rather it can lead to "method sterility". On the other hand, Gadamer also noted that for a researcher to have the ability to be a creative researcher, i.e. to have the ability to ask productive questions, a mastery of the methods used in research is required. An unhealthy overemphasis on either can lead to a lack of quality.

2.3 Qualitative versus quantitative research

As ten Have [18] suggests, quantitative versus qualitative is not a matter of rigorousness or quality, but *is* a matter of your focus and your research question:

> Quantitative research investigates . . . a relatively *small number of features* [that are] studied across a *considerable number of cases*. . . . Qualitative research, on the other hand, is especially used to study . . . a relatively *small number of cases* of which *many aspects* are taken into account. [18, p. 173, our italics]

ten Have's reminder can be translated into a choice between the size of your sample and the richness of your data (i.e. number of features, nuances and aspects studied). The goal in qualitative as well as quantitative research is "to produce theoretically structured descriptions of the empirical world that are both meaningful and useful" [19, p. 225] that in an dialogue between ideas and evidence produces "a representation of social life . . . presented along with the thinking that guided the construction of the representation" [20, p. 55].

However, qualitative research is often more difficult for engineers to become accustomed to than quantitative studies [cf. 21]. Quantitative research usually relies heavily on positivist traditions that assume an objective reality or on post-positivist traditions, which falsify a hypothesis generated by theory in the same way as found in many technical studies. Rarely do scientists and engineers discuss their underlying epistemology, nor necessarily know what it constitutes. The implication of this is that many researchers who move from technical to educational research, intent on ensuring 'rigour' and proving to their engineering peers that their work is valid, often adopt a positivistic stance that, in their own technical research, may not have been immediately apparent. This appearance of rigour, as opposed to quality, can tempt us to favour studies which adopt this epistemology.

> One of the dangers in conducting qualitative research is that it may appear easy and less rigorous than quantitative research. While quantitative research requires use of statistical methods which can provide an aura of trustworthiness, qualitative research can appear at first glance as if it simply involves interviewing a few people and then writing up a summary . . . In fact, qualitative research can be just as difficult to conceptualize, and be as methodologically and theoretical[ly] challenging, if not more challenging, than quantitative research. [9, p. 172]

This concern for work to be accepted as being rigorous often leads to an over-dependency on quantitative studies. What results is a fictional dichotomy between qualitative and quantitative studies, which is often confused with the underlying epistemology. As succinctly expressed above by Ragin [19, 20], the goal of research is to produce descriptions of the empirical world in a dialogue between *ideas* and *evidence*. This is why being explicit about the underlying epistemology and ontology is so critical. It is not possible to understand or judge the quality of work and the claims it makes unless its epistemology and ontology are understood.

3. Some quality criteria for research presented in the literature

As a backdrop to our discussion concerning quality, we have summarised some proposed, more "universal" quality criteria for chemistry, mathematics, EER and qualitative education research in table 1.

3.1 Criteria for quantitative as well as qualitative research

3.1.1 Quality criteria for mathematics education research (Niss)

In mathematics education research, quality criteria have been discussed extensively. One reason for this might be that this field initially had to struggle to be accepted as a legitimate branch of mathematics research [16]. Within this debate, Niss [22] has written a paper examining the question "what is quality in a PhD dissertation in mathematics education?" Niss divided his discussion into two main themes: *quality of the underlying research* and *quality of the dissertation as a report of the work done*.

The *quality of the underlying research* is further divided into the subcategories *quality of the research question(s)*, *quality of the research design* and *quality of the research findings*. Niss [22] considers the *quality of the research question*s posed to be the first important part of research quality. The primary characteristics of quality in this aspect are the questions of *clarity and precision*, but questions have to be genuine, non-rhetorical, of scholarly interest and have significance and relevance. Furthermore, the depth and the originality of a research question is another aspect of quality mentioned. Finally, he notes that a research question must be researchable. The *quality of the research*

design lies in the capacity of the design and methods to provide justifiable answers to questions in general and especially to the questions posed. Approaches that produce clear, strong and complete answers are of high quality. For Niss it is important that questions come first and design and methods come second. The fundamental criterion regarding *quality of the research findings* is that they add to scientific and scholarly knowledge and insights in the discipline. For Niss this means that findings are genuine, non-trivial, falsifiable and solid (trustworthy). Furthermore, the findings' applicability and relevance for mathematics education add to their quality.

It is worth noting that Niss [22] maintains that research designs and research findings should not be limited to being empirical, but can also be theoretical or a combination of the two approaches. This is contrary to the view of NRC presented in *Scientific Research in Education* [12] who limit research to questions ones that can be answered empirically.

The second main theme presented by Niss [22] is the *quality of the dissertation as a report of the work done*. This is further divided into the subcategories *the scientific/scholarly quality of the exposition contained in the dissertation* and *the communicative quality of the dissertation*. According to Niss, a good dissertation places the research done in a *scientific context* (i.e. relative to work done by others), presents a *clear* and *exhaustive* account of the research done, demonstrates a high degree of *balance* and *harmony* between different parts of the dissertation, *argues exhaustively* and *sufficiently* for the results obtained and offers a *thorough discussion* of the validity, reliability, strength, limitations, alternative interpretations and significance of the results. What Niss calls "*non-trivial trivialities*" such as referencing should be handled in an appropriate manner, i.e. it should be clear what are original ideas and what is borrowed from others. Furthermore, a high degree of *methodological* and *philosophical reflection* demonstrated in the dissertation contributes to a high quality. On the other hand, second- or third-hand accounts and quotations reduce the quality since there is a risk that the original authors' position is distorted or presented in an unbalanced way. The *communicative quality of a dissertation* is obtained by it having a *transparent structure* and a *clear* and *omnipresent "read thread"*. *Sharp definitions* of all key concepts should be provided.
Furthermore the dissertation should have a good *balance* between taking too much for granted and being over-pedantic or too repetitive; the *exposition razor* "when writing . . . strive to express yourself in the simplest possible way, without compromising conceptual clarity and exhaustive argumentation" [22, p. 17] should be applied.

In our minds, Niss [22] criteria are highly applicable to research papers in EER and not only dissertations. However, we have, in our brief summary, left out his discussion of the advantages and disadvantages of writing a dissertation as a monograph or as a collection of (published) papers together with an introduction. This discussion is outside the scope of this paper and this issue is also formally regulated in different ways in different countries and at different universities.

3.1.2 Quality criteria for chemistry education (Eybe and Schmidt)

Whilst Niss [22] arrived at his criteria through a combination of experience and theoretical reflection, Eybe and Schmidt [23] arrived at their criteria via an investigation into the criteria found in literature and an empirical investigation into the criteria *de facto* applied in 81 papers published in the *Journal of Research in Science Teaching (JRST)* and the *International Journal of Science Education (IJSE)*.

Eybe and Schmidt [23] found six main categories in their investigation, namely *theory-relatedness*, *the quality of the research question*, *methods*, *presentation and interpretation of results*, *implications for practice* and *competence in chemistry*. They also presented five papers that they considered met aspects of their criteria in an exemplary way.

The category *theory-relatedness* is described as having the criteria *the theory base* (i.e. making the theoretical underpinnings and basic assumptions explicit) and *reference to previous research* (i.e. establishing that the paper has a relationship to previous research).

Four criteria were found for the *quality of the research question*: *connection to existing literature* (e.g. in light of previous research, are specific research questions reasonable?), *relevance for practice* (e.g. it should be explained why research questions are relevant for teaching), *ethical issues* (i.e. in every phase of a study ethical issues should be considered) and research questions should present *hypotheses that can be falsified*. In regard to qualitative studies Eybe and Schmidt [23] clarify the criterion for falsifiability and note that such studies are more complex and that open questions are more common. However, they maintain that it is not scientifically appropriate to seek verification and thus it is questionable whether research should seek to verify the qualities of specific teaching methods.

For the *methods* criteria, Eybe and Schmidt [23] see a fundamental difference between criteria for quantitative and qualitative methods. Still, it is essential for both types of method to use approaches that are appropriate for the research question and the context investigated. For quantitative methods,

reliability (exactness of measurement), *validity* (does a test measure the intended variables?) and *level of significance* (statistical significance) are put forward as important criteria while the criteria for qualitative methods are *documentation of procedures* (data collection and analysis should be described in detail), *interpretation by logical interference* (conclusive logical flow in the interpretation and alternative interpretations should be discussed), *systematicity* (transparency regarding the rules followed, but there should also be a balance between being systematic and being flexible), *closeness to subjects* (preferably research should be conducted in natural settings and the researcher and the subjects should meet on equal terms), *communicative validity* (interpretations should be discussed with the subjects), and *triangulation* (use of different theoretical and methodological approaches to tackle research questions).

The remaining three main categories are presented without any further subcategories. High quality in *presentation and interpretation of results* is achieved by presenting results in sufficient detail, addressing limitations and incorporating alternative interpretations. Furthermore, statements should not be presented without supporting data, nor should data be presented that are not used to support statements. Quality in *implications for practice* requires research results that are relevant for educational practice. Finally Eybe and Schmidt [23] maintain that *competence in chemistry* is an important quality criteria.

3.1.3 Criteria for EER (Borrego and Bernhard)

Borrego and Bernhard [2] have described and discussed the emergence of EER as a field of inquiry. The aim of their review was to introduce EER *per se* and to describe U.S. and Northern and Central European approaches to it as examples of research traditions with slightly different views on which types of research questions are important and what constitutes research quality. As a way forward they proposed the following sets of criteria which they considered to reflect "broader conceptions of quality and encourage international interactions" [2, p. 37].

Quality scholarship in engineering education is:

- Inspired by real educational problems.
- Informed by theory and other literature describing prior work within and beyond the field/home country.
- Systematic and intentional, with documented decisions ideally based on well-planned collection and analysis of empirical data.
- Consistent with the perspectives and chosen methodologies (quantitative, qualitative or mixed).
- Presented (at least in part) in a form that engineering academic staff can understand and use, including by discussing implications of the research.
- Situated in international and interdisciplinary contexts, by demonstrating awareness of how common the problem is, what is being pursued elsewhere, and the likelihood that results are or are not generalizable/transferable to other contexts (disciplines and/or countries). We note that in order for an EER topic to be worthy of inquiry, it need not be broadly generalizable. [2, pp. 37–38]

Borrego and Bernhard [2] see their criteria as consistent with other lists of criteria, such as those, for example, offered by the *Journal of Engineering Education* and the *European Journal of Engineering Education*, but they argue that the "common ground is more clearly articulated".

The criteria presented in section 3.1 are quite general. Since the emphasis in this paper is on quality criteria for *qualitative* EER, in the following sections we present criteria specifically in relation to qualitative education research.

3.2 Criteria for qualitative research

Beside the criteria for qualitative education research that we describe in this section, it should be noted that there is an ongoing scientific debate related to qualitative *engineering* research. It is beyond the scope of this paper to review this discussion but Dittrich et al. [24] and Halldórsson and Aastrup [25] serve as examples of discussions related to quality criteria for *qualitative* engineering research.

3.2.1 Eight "Big-Tent" criteria for excellent qualitative research (Tracy)

Tracy [26] has presented and explored eight, "big-tent", criteria for quality in qualitative research. By using the term "big-tent" she aimed to underline that her proposed criteria encompassed different paradigms in qualitative research and that they are expansive and flexible. Her main criteria are *worthy topic*, *rich rigour*, *sincerity*, *credibility*, *resonance*, *significant contribution*, *ethics*, and *meaningful coherence*. Her criteria describe the *ends* (i.e. goal) separated from *means* (e.g. methods and practices) to reach these ends.

Worthy topic is achieved by the topic of research being relevant, timely, significant and interesting. *Rich rigour* is achieved by studies employing sufficient, abundant, appropriate, and complex theoretical constructs, sample(s), context(s), processes for data collection and analysis, spending enough time in the field to collect interesting and significant data,

and having enough data to support significant claims.

Sincerity is achieved by studies that are characterised by researchers being self-reflective as regards their subjective values, biases, and inclinations and that the studies are transparent as regards methods and challenges.

Credibility (i.e. the trustworthiness and plausibility of findings) is achieved in research that is marked by the use of in-depth descriptions, concrete detail, making explicit tacit knowledge, and showing rather than telling. Furthermore, the use of crystallisation or triangulation, multivocality and member reflections contribute to increased credibility.

Resonance (i.e. the ability to affect and reverberate meaningfully with an audience) is achieved through aesthetic and evocative writing and the use of representations, naturalistic generalisations and presentation of transferable findings.

The research provides a *significant contribution* (e.g. does the study extend knowledge? improve practice? etc.) by being theoretically significant (extending, building and critiquing disciplinary knowledge), heuristically significant (e.g. piques

Table 1. Some proposed quality criteria for education research

Quality criteria in Chemistry Education Research	**Quality criteria in Engineering Education Research**	**Quality criteria in Qualitative Education Research**
Eybe and Schmidt [23]	Borrego and Bernhard [2]	Tracy [26]
Theory-relatedness • The theory base • Reference to previous studies *The quality of the research question* • Connection to existing literature • Relevance for practice • Ethical issues • Falsification of hypotheses *Methods: Appropriateness of the method* *(a) Quantitative methods* • Reliability • Validity • Level of significance *(b) Qualitative methods* • Documentation of procedures • Interpretation by logical inference • Systematicity • Closeness to subjects • Communicative validity • Triangulation *Presentation and interpretation of results* *Implications for practice* *Competence in chemistry*	*Quality scholarship in engineering education is:* • Inspired by real educational problems. • Informed by theory and other literature describing prior work within and beyond the field/home country. • Systematic and intentional, with documented decisions ideally based on well-planned collection and analysis of empirical data. • Consistent with the perspectives and chosen methodologies (quantitative, qualitative or mixed). • Presented (at least in part) in a form that engineering academic staff can understand and use, including discussing implications of the research. • Situated in international and interdisciplinary contexts, demonstrating awareness of how common the problem is, what is being pursued elsewhere, and the likelihood that results are or are not generalisable/ transferable to other contexts (disciplines and/or countries). We note that in order for an EER topic to be worthy of inquiry, it need not be broadly generalisable.	*Worthy topic: The topic of the research is* • Relevant • Timely • Significant • Interesting *Rich rigour: The study uses sufficient, abundant, appropriate, and complex:* • Theoretical constructs • Data and time in the field • Sample(s) • Context(s) • Data collection and analysis processes *Sincerity: The study is characterised by:* • Self-reflexivity about subjective values, biases, and inclinations of the researcher(s) • Transparency about the methods and challenge *Credibility: The research is marked by:* • In-depth description • Triangulation or crystallisation • Multivocality • Member reflection *Resonance: The research influences, affects, or moves particular readers or a variety of audiences through:* • Aesthetic, evocative representation • Naturalistic generalisations • Transferable findings *Significant contribution: The research provides a significant contribution:* • Conceptually/theoretically • Practically • Morally • Methodologically • Heuristically *Ethical: The research considers:* procedural, situational and culturally specific, relational, and exiting ethics *Meaningful coherence: The study* • Achieves what it purports to be about • Uses methods and procedures that fit its stated goals • Meaningfully interconnects literature, research questions/ foci, findings, and interpretations with each other.
Quality criteria in Mathematics Education Research	**Quality criteria in Qualitative Education Research**	
Niss [22]	Larsson [27, 28]	
Quality of the underlying research • Quality of the research question(s) • Quality of the research design • Quality of the research findings *Quality of the dissertation as a report of the work undertaken* • Scientific/scholarly quality of the exposition, • Communicative quality.	*Quality of a study in general* • Perspective awareness • Internal consistency in a study • Ethical values *Quality of the results:* • Richness in meaning • Structure • Contribution to theory development and new knowledge *Validity of the results:* • Discourse criterion • Heuristic value • Empirical anchorage • Consistency • Pragmatic criterion.	

curiosity or initiates new research), practically significant, and/or by being methodologically significant (introducing and exemplifying new methodologies).

Research being *ethical* is not only a means but is also an end in itself according to Tracy [26]. A variety of practices attend to ethics in research such as procedural (categorical) ethics (e.g. general principles for ethics dictated by laws or regulations), situational ethics (e.g. the researcher should always reflect on the specific circumstances in a particular setting), relational ethics (e.g. the consequences of research on others), and exiting ethics (e.g. related to ethics after data collection is finished and how results are shared).

Finally, *meaningful coherence* is achieved by studies that deliver what they purport to be about, that use procedures and methods that fit their stated goals and that meaningfully interconnect literature, research questions or research foci, findings, and interpretation.

3.2.2 Quality in qualitative studies (Larsson)

The criteria for quality in qualitative studies put forward by Larsson [27] were first published in 1993 and were re-published in 2005 as a part of a selection of the best papers from the 25-year history of the journal *Nordisk pedagogik* [Nordic Pedagogy]. As such, the work has been influential in the Nordic countries and, for example, the recent thesis by Haglund [29] makes direct reference to Larsson's criteria. However, the paper by [27] is written in Swedish and is only available in English within a set of conference proceedings [28]. Consequently, these criteria are relatively unknown outside the Nordic countries.

In our opinion, the criteria presented by Larsson [27, 28] offer a good structure and we have taken them as a point of departure in the our development of criteria. The rationale for this selection is presented more extensively below in the introduction to section 4. To avoid unnecessary repetition, here we only very briefly mention Larsson's criteria in their original form. The criteria are explained in more detail in sections 4.2–4.4.

Larsson's criteria [27, 28] are divided into three main themes: *Quality of a study in general* with subcriteria *perspective awareness*, *internal consistency within a study*, and *ethical values*; *Quality of the results* with subcriteria *richness in meaning*, *structure*, and *contribution to theory development and new knowledge*; and *Validity of the results* with subcriteria *discourse criterion*, *heuristic value*, *empirical anchorage*, *consistency*, and *pragmatic criterion*.

Larsson clearly states that there is some tension between the criteria and it is not possible to fulfil them all completely; these tensions are handled differently in different research approaches and certain criteria are given a stronger or weaker emphasis depending on the approach taken.

3.2.3 Criteria for interpretive EER (Walther et al.)

In parallel to our work, Walther and Sochacka [5] have recently published a valuable contribution to the discussion of quality in interpretive engineering education research (i.e. a subset of qualitative research in which knowledge is gained through investigation into and interpretation of individuals' or groups' lived experience). They have applied the metaphor of quality management from engineering and have especially focused on process quality, i.e. generating and handling data in research. The focus is exclusively on validation and reliability, since they are seen "as fundamental ways to contribute to and demonstrate the overall quality of interpretive research" [5, p. 639]. In our view, discussion of quality cannot be limited only to *validity* and *reliability* since there are other aspects that are important contributors to the quality of a study.

Walther et al. [5] present five constructs related to managing quality in interpretive EER, namely *theoretical validation* (the fit between the theory generated and the reality investigated), *procedural validation* (is research designed to improve the fit between theory and reality?), *communicative validation* (concerns the co-construction of knowledge in the investigated social context as well as within the relevant research communities), *pragmatic validation* (are theoretical constructs compatible with empirical reality?), and *process reliability* (how are random influences in the research process limited?). In later versions *ethical validation* has been added as a sixth construct [30].

4. Tentative quality criteria for (qualitative) EER

Our aim with this paper is not to present a complete review of the education literature discussing quality criteria for research, but to present tentative criteria for qualitative research in EER as a contribution to an ongoing debate. The criteria we propose are ambitious in the sense that we see them as universal criteria for quality in qualitative engineering education research. By universal we mean that the criteria are intended to be applicable to most aspects of research quality (and not, for example, limited to validity and reliability as is often the case in the literature) and that they would be relevant for most qualitative research approaches.

The criteria for quality that we propose are drawn from criteria outlined by Borrego and Bernhard [2] as well as Larsson's [27, 28] criteria and those

suggested by Eybe and Schmidt [23], Niss [22] and by Tracy [26]. The reviewed criteria are summarised in Table 1 and our tentative criteria are summarised in Table 2. We maintain that there are sometimes contradictions or tensions between criteria and an adequate balance has to be found, and that, depending on the type of study, certain criteria are more important than others.

Criteria for quality cannot be proven right or wrong (cf. a similar statement regarding theories and methodologies in section 4.2.1 below). However they can be judged with respect to their consequences and their aesthetic qualities (cf. Sandelowski [31]). We find the structure—*quality of a study in general—quality of the results—validity of the results*—of Larsson's criteria very useful. This structure is not as visible in other proposed criteria. Furthermore, the criteria under the heading *Quality of a study in general* (section 4.2) are valid for quantitative research as well and most criteria under the heading *Quality of the results* are applicable for quantitative research (although sometimes with a slightly different interpretation). Thus, most of the criteria are valuable for quantitative research and the impact of the criteria can extend beyond qualitative EER-studies. Furthermore, we consider the criteria to be applicable to empirical as well to theoretical studies. In our experience, a discussion regarding research quality is much needed in the EER-community in relation to quantitative as well as qualitative approaches.

To Larsson's original criteria [27, 28] under the heading *quality of a study in general* we have added the criteria *acknowledging different knowledge traditions and cultures* [2], *being informed by theory and other literature describing prior work* [2, 22, 23], *research questions* [22, 23, 26], and *research design and quality of empirical data* [2, 22, 23, 26]. Under the heading *quality of the results* we have included the criterion *presentation of results* [2, 22, 23]. Most of these criteria are implicit in Larsson's other criteria [27, 28], but from our experience from involvement in the EER-community we consider it important that these criteria should be explicit. In our view, these criteria fit with the original criteria proposed by Larsson.

The falsifiability criterion put forward by Eybe and Schmidt [23] and Niss [22] is incompatible with our view; this criterion is based on Popper's view of what constitutes a science [32]. This view is contested in the theory of science [e.g., 33], but it is beyond the scope of this paper to review that debate. However, we note that Popper's view is not in line with the practical epistemology used in engineering [34, 35] and that it would exclude most qualitative research as well. Furthermore, we have not included a parallel to Eybe and Schmidt [23] criterion *competence in chemistry*, i.e. *competence in engineering*. Although we would argue that it is important that engineering competence is present in the EER-community as whole and it would probably contribute to the quality of research, we maintain that many other competencies are needed in EER and exactly which competencies are needed in an *individual* study would depend on the research question and the methodologies used. Therefore, we find such a criterion unnecessarily exclusive and potentially detrimental to research quality in some cases.

In our paper we put forward *Perspective awareness* (section 4.2.1) and *Acknowledging different knowledge traditions and cultures* (section 4.2.2) as important quality criteria. However, these are not easy aspects to understand when coming from a positivist position which maintains that there is only one world and what we have to do is to uncover it. Different perspectives have little meaning in this epistemology except by way of explaining how a student or researcher might relate the truth being explored to their own experiences and thereby to understand it. It is more usually discussed in engineering education when indigenous knowledge is considered (see, for example, Hess and Strobel [36]) but it is just as relevant for all studies. For example, the first author's epistemological position [37] is stated below:

> *I am most importantly influenced by pragmatism [38-42], post-phenomenology (a synthesis of phenomenology and pragmatism within the philosophy of technology developed by Don Ihde [43]), phenomenography [44, 45] and by ethnomethodology [46]. To be extremely brief, these different approaches differ in that they investigate different aspects of the life-world but share a mainly common view on human experience, an inter-relational non-dualistic ontology and an interest in praxis and abilities [47]. One implication of this epistemological position is that I subscribe to Rorty's [48] statement that "[we should not] view knowledge as a matter of getting reality right, but as a matter of acquiring habits of action for coping with reality"* (cf. Peirce [38]).

This perspective and understanding of what knowledge *is* and *can be* is subtly different from epistemology and yet also influences the decisions made in choice of methods, research questions and analysis, as well as writing traditions. In order to understand quality, we would ideally need to know something more about the position of the author(s) in this regard.

In the next section we provide a hypothetical study to exemplify our criteria further.

4.1 Our hypothetical study

Problem statement: Delia Harris wanted to study her students' conceptual understanding of first year level materials science concepts, after having taught them using an interactive approach. The course

focussed on the thresholds of understanding necessary for future study, which Delia hoped would transform their ways of thinking about materials properties and structure and yet which her students found troublesome. In each of the criteria discussed below we will demonstrate how Delia's study could be framed in order to address the quality criteria. We have chosen to use a hypothetical example so as not to suggest that one approach is better than another, but only how one particular study addresses the criteria. It is also possible in our hypothetical case to show all criteria but it is very unlikely that all quality criteria will need to be applied to all work. However, if criteria are missing, the authors (and reviewers) need to know that this is because they are not needed, rather than that they are simply ignorant of the need. Finally, with a hypothetical situation we can look at the process of how the research was developed, rather than only at the paper 'product'. This will be of value when readers are considering setting up a research grant proposal, as well as in considering how to frame their studies and present their work at the end.

4.2 Quality of a study in general

The criteria in this section are related to the quality of the study as a whole. They are, for example, related to how the interplay between theory, epistemology, methods, methodologies and research designs are handled in a study [10].

4.2.1 Perspective awareness

> *Without theory there is nothing to research* . . . In this way, theories provide both a framework for critically understanding phenomena [and] a basis for considering how what is unknown might be organized . . . However, as used here, models, concepts and theories are self-confirming in the sense that *they instruct us to look at phenomena in particular ways. This means that they can never be disproved but can only be found to be more or less useful* . . . A methodology refers to the choices we make about cases to study, methods of data gathering, forms of data analysis etc. in planning and executing a research study. So our methodology defines how we will go about studying any phenomenon . . . *Like theories, methodologies cannot be true or false, only more or less useful*. [49, pp. 3–4, our italics]

We suggest that the theoretical framing provides a 'lens' through which we see our subject/object of study. Learning theories will help frame what we mean by good quality learning, research theories will help us decide what constitutes good quality research and what question/aspects of learning we might focus on. As proposed in the above quote, the bias of the authors and their perspective and positioning theoretically and methodologically need to

Table 2. Tentative quality criteria for (qualitative) engineering education research

§ 4.2 Quality of a study in general	**§ 4.3 Quality of the results**	**§ 4.4 Validity of the results**
• *§ 4.2.1 Perspective awareness* All studies have a perspective or 'lens' through which the researchers have chosen to view their subject • *§ 4.2.2 Acknowledging different knowledge traditions and cultures* Respect and awareness of the perspective of other researchers • *§ 4.2.3 Upholding ethical values* How can the study contribute to enhancing the human condition? • *§ 4.2.4 Informed by theory and other literature describing prior work* A researcher cannot perform significant research without first understanding the literature in the field. • *§ 4.2.5 Research question* Worthy topic: Relevant, timely, significant, interesting • *§ 4.2.6 Internal consistency in a study* Consistency between research questions, methodology, epistemology, ontology and data analysis • *§ 4.2.7 Research design and quality of empirical data* (where appropriate) Systematic and intentional, with documented decisions ideally based on well-planned collection and analysis of empirical data.	• *§ 4.3.1 Richness in meaning* Capture the essentials and at the same time maintain the nuances. Highlight what is unique to the specific phenomenon at hand. • *§ 4.3.1 Structure* Interpretations should be well-structured. It should be possible to follow the reasoning. • *§ 4.3.2 Contribution to theory development and new knowledge* How well do the results relate to earlier theory? What is the original contribution of the study; something decisive or just a note in the margin? • *§ 4.3.3 Presentation of results* Presentation relevant to proposed audience, clear and precise language, and a good balance between different parts of the presentation.	• *§ 4.4.1 Discourse criterion* In this criterion, validity is viewed as a conversation about the world (or reality). The quality and exposition of argumentation and interplay of meanings. • *§ 4.4.2 Heuristic value* To what extent will a reader be convinced by the presentation of the study in seeing a particular aspect of reality in a new way? • *§ 4.4.3 Empirical anchorage* The relationship between 'reality' and interpretation • *§ 4.4.4 Consistency* The interplay between part and whole • *§ 4.4.5 Pragmatic criterion* Consequences of what the results brought about (for example in relation to teaching). This is an aspect of what also is called ecological validity.

be made explicit, as interpretation of data will be based on this. In order to appreciate the communicated results and their validity, the reader also needs to "see through the lens" of the authors.

Delia has a background in engineering, but has worked for several years with phenomenographers and has an appreciation of the ways in which they view learning. Phenomenography as an epistemology or approach to understanding ways of knowing, and as a learning theory seems to fit with her beliefs about what learning is: negotiated between teacher and student, that the ultimate learning experience must change you as a person, and that the world will therefore in some way also change. Recently she has come across threshold concept theory, and although this seems less well developed, she finds it very helpful to frame her curriculum development and teaching. She positions herself with the idea that teaching is facilitation and that her role is to engage the students in a dynamic whereby they might enter the 'liminal space' of learning and ***cross*** *the threshold so that new ways of knowledge open up for them. Delia describes all of this in her paper before setting out on her story of the research she has done.*

4.2.2 Acknowledging different knowledge traditions and cultures

> No research paradigm has a monopoly on quality. None can deliver promising outcomes with certainty. None have the ground for saying 'this is it' about their designs, procedures, and anticipated outcomes. [50, p. 36]

High quality with respect to this criterion is to display an awareness of, and respect for, *other* researchers' perspectives, and to understand a piece of work within its own theoretical framework. This criterion is critically important for anyone involved in reviewing or editing other's work. As quality is negotiated by the community within this category, it will be possible for more appropriate matching of reviewers to work by Journal Editors. It is especially important for reviewers to acknowledge their ignorance of others' traditions and to decline to review studies not within their area of expertise.

In her paper, Delia acknowledges that her choice of research method and theoretical lens are not the only ones which would be useful but, given her own background, the context of the study and the nature of the teaching and learning to be studied, she believes they are the most appropriate.

4.2.3 Upholding ethical values

> Research quality . . . in a broad sense is a holistic judgement, from which *no individual requirement could be removed* . . . Quality is judged by the compounded qualities originality, external and internal validity, precision and ethics. *This means that the requirement for good research ethics is also included*. [51, p. 24, our translation and italics]

While ethical considerations are very briefly discussed in the National Research Council [12, p. 93] report and in parts even described as something that "may weaken the strength of [a] research [design]", the Swedish Research Council maintains that good research ethics are essential for high quality, as demonstrated in the quote above. Furthermore, the Swedish Research Council [51] proposes the following eight 'commandments':

1. You should speak the truth about your research;
2. You should consciously scrutinise and describe the basis for your studies;
3. You should openly describe methods and results;
4. You should openly account for commercial connections and other liaisons;
5. You should not steal results of research from others;
6. You should keep good order in your research, for example through documentation and keeping archives;
7. You should aim to conduct your research without harming humans, animals or the environment;
8. You should be fair in your judgement of other people's research

The above is important because they maintain that ethics is much more than simply conducting research "without harming humans or animals". Fraud, withholding contradictory data, abuse of intellectual property, plagiarism, or lack of precision, in for example superficial studies about intelligence, criminality, ethnicity and so forth are all seen as examples of poor research ethics.

Delia acknowledges in her study that, as she wishes to study her own students learning, and that, as she has an ownership of the ideas for both the pedagogical innovation studied and the research, she clearly has a bias towards seeing the positive in her students' learning outcomes. She chooses to partner with two other researchers who conduct the interviews and analysis, and who work with Delia to write the paper. One is a research associate appointed to work on the research and the other has an educational research background. She makes the roles of each explicit in the paper. [It should be noted that in this paper we refer to 'Delia' in the first person, and include her co-authors within this].

4.2.4 Informed by theory and other literature describing prior work within and beyond the field/ home country

> A substantive, thorough, sophisticated literature review is a *pre-condition* for doing substantive, thorough, sophisticated research. *'Good' research is good*

> *because it advances our collective understanding.* To advance our collective understanding, a researcher or scholar needs to understand what has been done before, the strength and weaknesses of existing studies, and what they might mean. *A researcher cannot perform significant research without first understanding the literature in the field.* [52, p. 3, our italics]

A thorough review of previous literature serves as a foundation for any research project by placing it in the broad historical context of previous studies both nationally and internationally. Furthermore, "a good literature review is the basis of both theoretical and methodological sophistication, thereby *improving the quality and usefulness of subsequent research*". [52, p. 4, our italics]

Delia has an extensive literature review section in her paper. As she did not have time to locate all relevant papers herself, she used some of her research allowance to hire a research assistant to conduct a literature study for her, on the theoretical background and the context of study. The RA created a spreadsheet which clearly indicated the relevance and importance of each paper and Delia was then able to study the most relevant papers and include pertinent aspects of these in her report. The research assistant was also one of the co-authors mentioned above—so questions of intellectual property and appropriate authorship were addressed.

4.2.5 Research question

> A crucial but typically undervalued aspect of successful scientific investigation is the quality of the question posed. [12, p. 55]

For example, Niss [22] maintains that "the issue of quality of the *research* undertaken takes its point of departure in research question(s) posed". The questions' *clarity*, *precision*, if they are *genuine* and *non-rhetorical*, their scientific and scholarly *interest*, if they are *significant*, *relevant*, *original* and *researchable* are seen as important aspects of quality. It is often argued "that research questions should drive data collection techniques and analysis rather than vice-versa" [53, p. 16]. However, Larsson [27] argues this is too simplistic since the choice of methods and approaches also includes choice of perspectives that will influence the kind of results obtained. In a similar vein, it is reiterated that research "is always conducted within some broader understanding of what *constitutes legitimate inquiry* and *warrantable knowledge*" [54, p. 98, our italics]. This is expressed by Case and Light [8, p. 189] when they write "one's choice of methodology will constrain what questions one can ask". The *explicit* formulations of research questions do not play a prominent role in some disciplines and research traditions while in others they are formulated *post hoc*. Nevertheless, we consider that all research attempts to answer research questions, even if they are implicit. Good research is about asking good questions. Even if the *explicit* formulation of the research question is *post hoc* it serves the function of being a "clear signpost [without which] the readers will be lost . . . not know[ing] the central ideas addressed in [the] study" [55, p. 112].

Delia's two co-authors helped her frame the research question for the study. She came to them with a rough problem brief but no specific problem statement. They assisted her by questioning assumptions she was making, and by probing her meanings and intentions. Eventually the right research question was formulated. She asked: 'does the interactive approach I have adopted in my classroom assist in students' conceptual development of core threshold concepts in first year materials science?'

4.2.6 Internal consistency in a study (including epistemology, ontology with methodology)

> Harmony should exist between the research question, assumptions about the research and the nature of the phenomenon to be studied, data collection, and methods of analysis. [27, p. 21]

This is perhaps [27] the most commonly used criterion, but as mentioned previously, one that is often missing [8, 9]. Beneath the surface of this criterion is the idea of *construction of a whole* where single parts are integrated and contribute to the building of a single piece of research. It requires much self-reflection about 'what I think knowledge means'—both the developing knowledge of the learner and how we develop knowledge as researchers.

After Delia had framed her question, knowing her perspective on learning and her chosen theories, she then worked to find the most appropriate research methods. How could she actually find out whether the interactive approach she was using was related to student learning outcomes? Could she, in fact, find out what the learning outcomes were? Was she supposed to be comparing this with previous years? But if she did that it would be with different students. . .. Delia became very confused, and together with her colleagues, reframed her question once again. Now she asked 'are students in my classroom entering a threshold, or some kind of liminal space of learning with respect to two selected concepts in materials science (which for the purpose of this study are assumed to represent a threshold)?' and if they are 'how are they passing through and what assists them in doing so?' She realised that she needed to explain her assumptions in her paper, with respect to previous work but she was now much happier with her question. She and her colleagues decided that she would use an adapted approach combining phenomenography and threshold concept theory—conducting and analysing semi

structured interviews for different stages of understanding as the learners pass through the liminal space.

4.2.7 Research design and quality of data

> The capacity of the [research] design and methods to provide justifiable *answers to questions* in general . . . and in particular justifiable answers to *those questions* that have been posed for the research endeavour at issue, is a primary quality parameter. [22, p. 11, italics in original]

The research design is the overarching plan that a researcher creates and uses to guide the intended research in order to come up with *justifiable answers* to the *research questions* posed [22, 55]. This means that the research design guides the choice of methods and hence the type and quality of data collected. It is essential, as is pointed out in section 4.2.6 that there is a "harmony" and good interplay between the research question, theoretical assumptions (such as epistemological and ontological assumptions), prior research in the field (thorough reading of the literature is important) and ethical values. The quality of a research design is, thus, intimately related to internal consistency. If the study is empirical, there is a need for rigorous, systematic and intentional data collection and analysis, with documented decisions laid out in the paper. However, there *can* be conflict between rigour and flexibility (i.e. letting preliminary findings in a study for example guide subsequent research and data collection). The best balance between rigour and flexibility depends on the research question and on the nature of the data collected.

Delia created a plan based on all the decisions she had made so far and this included how many people she would interview, who they would be, how she would invite them, how she would analyse the data etc. She followed the steps for rigorous data collection and analysis set by phenomenographers as threshold concept theorists have yet to create any standards, but she made sure that what she did made sense within the whole framework of learning and knowing that she was working within. She explained all this in the paper.

4.3 Quality of results

In this section, criteria addressing the quality of the results and the descriptions and presentation of the results are discussed. In qualitative research, the results themselves and their presentation are usually intimately intertwined. Hence the criterion *presentation of results* is included in this section.

4.3.1 Richness in meaning and Structure

> Qualitative and quantitative research procedures are but different forms of the analytic practice of *re-representation* in science, in that both seek to *arrange and rearrange the complexities of 'raw' data*. [They] *differ . . . in their approach to re-representing complexity*. [The] very strength [of quantification and numbers]—that of simplification—is also its Achilles' heel! [54, p. 42, our italics]

The quotation above highlights the tension between the two different criteria *richness in meaning* versus *structure*. In qualitative research, it is essential not to lose sight of the complexities, and hence, *richness in meaning* is important. However, even in such studies there is a tension between *richness* and *structure* since an overview may be sacrificed if descriptions are "too rich". To ensure good quality as regards the criteria *richness in meaning* it is necessary that the essentials are captured and at the same time the nuances are maintained. It is important to highlight what is unique to the specific phenomenon at hand. For the criterion *structure* it is important that interpretations should have a good structure and it is possible to follow the reasoning. We would note that many quantitative studies in education have low quality as regards *richness in meaning* since, for example, settings and interventions are often poorly described and hence it is very difficult for a reader to understand what was done. Erickson and Gutierrez [56] remind us that a "logically and empirically prior question in response to 'Did it work?' is 'What was the *it*?'"

Delia decided that the richness in her case would relate to the actual concepts that students were learning. She had read several papers using phenomenography which she felt did not go into enough detail about the concepts which were being learned. She liked the way that threshold concepts studies were often co-authored by a subject specialist so that the complexities of the technical concepts were discussed. As a result, the interviews, the analysis and the paper take on a very technical flavour. The paper may not be well understood by those without the necessary technical background, but she felt it was important to demonstrate accurately what was happening in the liminal space of concept development. The paper then had to be well crafted so that it was meaningful to non specialists, even if not at the same level. The 'story' of the pathway through the liminal space needed to be laid out clearly and this took many iterations of data analysis and conversations between the co-authors as well as redrafts of the paper.

4.3.2 Contribution to theory development and new knowledge

The quality of the results with respect to this criterion is somewhat related to the quality of the literature review and the quality of the research, as described above. The researcher must have a good understanding of the field and good quality is related to the original contribution of the study to

our collective understanding; is it something decisive or just a note in the margin? How well are the results placed in context and related to earlier theory and knowledge?

Poor quality research tends to focus on reliability, by producing many repeated test results, at the expense of any meaningful result. We need always to ask ourselves whether the research is actually telling us anything meaningful, and if it is an honest reflection of the aims of the study.

Delia's study clearly takes the use of threshold concept theory as well as phenomenography further by combining the two in ways which are fruitful and transferable. She also learns much about her students' positions with respect to the liminal space which she places them in within her classes. This enables her to plan developments for the future and to share lessons learnt with other teachers.

4.3.3 Presentation of results

If earlier criteria can be seen as related to the scientific and scholarly quality of a study, this criterion is related its communicative quality. However, it should be noted that, for qualitative studies, the scientific and scholarly quality of results cannot be separated from the presentation of results as is typically the case for quantitative studies. Nevertheless, we see a value of having a separate criterion for presentation of results. The presentation should be relevant to the intended audience, language should be clear and precise and there should be a good balance between different parts of the presentation. This is not only a matter of rhetorical quality with a more fluent language free from grammatical errors but it also enhances the precision and clarity of the presentation. However, not only is the communicative/rhetorical quality important but what is sometimes called *akribeia* (from Greek ἀκρίβεια; precision, carefulness) is also essential; it is a matter of getting all the non-trivial trivialities such as, for example, citing and referencing, right. Interdisciplinarity is a notable issue here. Communicating across disciplines is not straightforward. We, as journal editors, note that when engineering educators attempt to navigate into other disciplines, such as social science, many papers are rejected by reviewers who are social scientists and ignorant of engineering communication styles, just as engineers are ignorant of social science methodologies. Studies of high quality are sometimes rejected from journals because they are not composed in the same style of narrative that is expected by some disciplines. For example, even presenting 'methods' is considered positivist by some scholars, such as strict Foucauldian discourse analysts, and hence seems a contradiction of epistemologies.

Included in this criterion is the issue that the instructions for authors for the specific journal or the specific conference should be adhered to. As this usually also means limits about how long a paper can be (and for conferences sometimes very strict limits) this means that there is a tension here with other criteria or even with the criteria of communicative clarity within this criterion. Even when there is no strict limit in length, the *exposition razor* put forward by Niss [22, p. 17] is an important tool and an author has to find a balance between length (with the intended audience in mind) and conceptual clarity and exhaustive argumentation.

Delia's paper is written in a very innovative way, as it presents very technical content, using complex adapted educational theories and methodology. She tells a story of the pathway of learning through liminal space. None of this may be familiar to the reviewer or the reader. Hence she has to spend a lot of time explaining the structure of the paper, clearly laying out the research steps and finally summarising key points of learning in technical and educational, research and teaching domains. This will be framed by her clear presentation of the ways in which her epistemology, methodology, methods, research questions and data analysis are connected, and how her personal bias colours the research.

4.4 Validity of results

Hitherto we have dealt with qualities of studies, reports and of the result characteristics. We shall now discuss the validity of results. However, in qualitative studies this issue is not straightforward and in some approaches the distinction between *quality of results* and *validity of results* may not be clear and free from complications. In some literature the term *trustworthiness* [e.g., 57] is used instead of *validity*, since the latter term is thought to bear "positivist" connotations.

4.4.1 Discourse criterion

Validity in this criterion is viewed as a conversation about the world and reality; validity is related to the quality of argumentation and the interplay of meanings and whether the findings of a study are able to withstand careful scrutiny and convince the scientific community. Complications arise when the dominant discourse, or 'common sense' is questioned by the authors, but not by the readers/reviewers, if reviewing within the dominant discourse, then counter hegemonic arguments would not withstand the scrutiny of 'rational argument'.

Delia is crossing disciplinary worlds in her research, so will have to concern herself with what constitutes validity within engineering and within education. What is 'common sense' to technical reviewers, who might understand the conceptual

understandings presented in the paper, will rarely include the methodologies of this paper. What is 'common sense' to educational researchers may not include such a discussion on the technical. Hence Delia has to prove that this paper is a valid piece of research—even though it contravenes rules from each domain. To attempt to address this she would explain everything very carefully including her methods, however some research traditions are very negative to explaining methods as this appears to be reductionist, as mentioned above. This requires us to educate our reviewers better as this aspect is not within the control of the author.

4.4.2 Heuristic value

In this criterion we are concerned about the extent to which we might see something in a new way. This criterion is often seen as an essential validity criterion in qualitative studies and is related to the significance of the study, or to the extent to which the research develops a conceptual framework which brings new ideas and thought processes to the literature, worthy in themselves of further work and thought, and how the reader is convinced. A successful gestalt of the results has rhetorical qualities.

This criterion will not apply to all studies. In Delia's case the reader may begin to see that learning involves confusion and that this is a necessary step towards crossing a threshold. This will bring into question the whole area of student evaluation and satisfaction surveys—no-one much likes to feel confused. Again this will appear in the outcomes and lessons.

4.4.3 Empirical anchoring

This criterion concerns the relationship between interpretation and 'reality' and it has relevance for empirical research in most paradigms in one way or another. This will vary hugely between different epistemologies: in some there is a one-to-one correspondence between reality and interpretation, i.e. only one interpretation is possible while in others multiple interpretations are possible depending on, for example, perspectives making it is possible to talk about "multiple realities". This again stresses the importance of disclosing the underlying epistemology—as notions of interpretation are clearly more complex when it is acknowledged that subjectivity is always present. For this reason, we use the term "empirical anchoring" and not "correspondence".

This criterion demonstrates the problems that might arise when engineers read Delia's paper, which we hope they will. They may have a hard time believing the data, if it comes from a different set of frameworks and epistemologies than they are used to. Even if within their own work they are used to using qualitative data, in this case they might expect statistics and a high sample number to 'prove' what has been claimed. Apart from the necessity for clearly laying out the methodology and analysis as discussed above, in Delia's case, this criterion speaks more to the need for appropriate selection and education of reviewers, rather than something which can be addressed by the authors.

4.4.4 Consistency

This criterion is central to the meaning of hermeneutics: an interpretation is built up by an interplay between the whole and its parts. The meaning of the whole is created from the parts and at the same time the meaning of each part depends on the meaning of the whole. Hence, there should be an interplay between parts and the whole in the results. The metaphor of a jigsaw puzzle is often used to illustrate this concept—each piece of the puzzle has to fit. This is even true if there are multiple realities being represented, as outlined in the previous section. If the underlying knowledge system being explored, the research approach and the research question are consistent with there being a complex system of integrated and inter-subjective experiences in the world, then many different views of the world may emerge which make up such a 'whole'.

Through all of the explanations of her own positioning, her choice of research methods, her understanding of what knowledge is, her adoption of threshold concept theory with its multiple liminal pathways, her framing of phenomenography with its notion of non dualistic collective learning, Delia is able to demonstrate the consistency within her results and with her thinking and of her studies, with their theoretical and epistemological underpinnings.

4.4.5 Pragmatic criterion

This criterion is related to the consequences of the results (for example in relation to teaching) and this is sometimes referred as "ecological" validity or the external value of a study. Engineers should be familiar with this criterion since it is related to the extent to which findings or a theory work in the "real world" and in education it could, for example, be related to the extent to which teaching based on the findings of a study result in better understanding of a certain content or better skills in some field.

Delia's study gives her direct input to improve her teaching for next year, as she now knows which concepts are misunderstood and partially understood, what it looks like to be confused and how to assess the students to show movement through the liminal space. She also has some insights as to which of her teaching methods worked to support this movement. This is all

laid out in the paper so others can also learn from the study.

5. Discussion

In this paper we present a tentative list of quality criteria for EER, with arguments, comparisons and examples. We initiated this development of quality criteria for qualitative EER some years ago because we felt a need to discuss quality issues within the EER-community in general and especially with regard to qualitative research. Marton [45] suggests that the most import aim of education is to prepare students for an unknown future and research as learning on a "collective level" [45, p. 154]. We decided to open up and invite you into a liminal space of learning about quality in EER.

Leading scholars such as Bochner [58] present *Criteria against ourselves* and Schwandt [59] argue for the *Farewell to criteriology*. Others, like Creswell [60], argue that the variety of research approaches in qualitative research requires a similar variety in criteria for quality. We consider that the values and the risks of criteria depend on how you read and understand them, but that we need to begin somewhere. Furthermore, as Cronbach and Suppes remind us, there will always be excellent studies that do not fit, at any given time, into *existing* criteria:

> Disciplined inquiry does not necessarily follow well established, formal procedures: Some of the most excellent inquiry is free ranging and speculative in its initial stages, trying what might seem to be bizarre combinations of ideas and procedures, or restlessly casting [about] for ideas. [61, p. 16]

Therefore, we do not see our criteria as fixed and eternal; instead they should be open for discussion, re-interpretation and revision.

A limitation of this paper is that it has not been possible for us to explore our criteria in depth, due to length considerations, and hence to discuss all aspects extensively. In a future paper we plan to discuss our criteria in more detail and to explore the tensions between different criteria.

6. Conclusion

A thorough discussion of quality criteria is important in the EER-community. Without an open and transparent discussion there is risk for hidden, random or subjective criteria. Furthermore, discussion of quality of research should not be confined to only a discussion about validity and reliability. Our suggested criteria are more encompassing and are grouped in three main themes: *Quality of a study in general*, *Quality of the results*, and *Validity of the results*. Although the focus in this paper is on qualitative research, we argue that the criteria under the first two headings *Quality of a study in general* and *Quality of the results* are important and mostly applicable for quantitative research as well. We do not see our proposed criteria as set; on the contrary criteria should be open for debate and revision to be fruitful guidance for future research in EER. Hence, it is our hope that our limited beginnings will promote thoughts that will contribute to a future discussion.

Acknowledgements—Earlier versions of this paper have been presented at the Nordic Network for Engineering Education Research (NNEER) seminar in Aalborg, Denmark, in 2012, at the SEFI annual conference in Thessanoliki [3], Greece, in 2012, and at the Research in Engineering Education Symposium (REES) in Kuala Lumpur [4], Malaysia, in 2013. We are grateful for the suggestions and feedback we have received at these events. We are, furthermore, grateful for comments and suggestions from three anonymous reviewers.

References

1. C. Baillie and J. Bernhard, Educational research impacting engineering education, *European Journal of Engineering Education*, **34**(4), 2009, pp. 291–294.
2. M. Borrego and J. Bernhard, The emergence of engineering education research as a globally connected field of inquiry, *Journal of Engineering Education*, **100**(1), 2011, pp. 14–47.
3. J. Bernhard and C. Baillie, Standards for quality of research in engineering education: A prolegomenon, in *proceedings of SEFI annual conference*, Thessanoliki, 2012.
4. J. Bernhard and C. Baillie, Standards for quality of research in engineering education, in *proceedings of Research in Engineering Education Symposium (REES)*, Kuala Lumpur, 2013.
5. J. Walther, N. W. Sochacka and N. N. Kellam, Quality in Interpretive Engineering Education Research: Reflections on an Example Study, *Journal of Engineering Education*, **102**(4), 2013, pp. 626–659.
6. C. Willig, *Introducing qualitative research in psychology: Adventures in theory and method*, 2nd ed., Open University Press, Buckingham, 2008.
7. M. C. Shelley II, L. D. Yore and B. Hand, Education research meets the "Gold Standard", in M.C. Shelley II, L. D. Yore and B. Hand (eds.), *Quality research in literacy and science education: International perspectives and Gold Standards*, Springer, 2009, pp. 3–15.
8. J. Case and G. Light, Emerging methodologies in engineering education research, *Journal of Engineering Education*, **100**(1), 2011, pp. 186–210.
9. M. Koro-Ljungberg and E. P. Douglas, State of qualitative research in engineering education: Meta-analysis of JEE articles, 2005–2006, *Journal of Engineering Education*, **97**(2), 2008, pp. 163–175.
10. C. Baillie and E. P. Douglas, Confusions and Conventions: Qualitative Research in Engineering Education, *Journal of Engineering Education*, **103**(1), 2014, pp. 1–7.
11. R. A. Streveler and K. A. Smith, Conducting rigorous research in engineering education, *Journal of Engineering Education*, **95**(2), 2006, pp. 103–105.
12. National Research Council, *Scientific research in education*, National Academy Press, Washington, DC, 2002.
13. S. T. Hopmann and K. Riquarts, Didaktik and/or curriculum: Basic problems of comparative didaktik, in S.T. Hopmann and K. Riquarts (eds.), *Didaktik and/or curriculum*, Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel (IPN), Kiel, 1995, pp. 9–40.
14. I. Westbury, Teaching as a reflective practice: What might didaktik teach curriculum?, in I. Westbury, S. Hopmann, and K. Riquarts (eds.), *Teaching as a reflective practice: The German didaktik tradition*, Lawrence Erlbaum, Mahwah, NJ, 2000, pp. 15–39.

15. L. B. Henriksen, Engineers and bildung, in J. Christensen, L. B. Henriksen, and A. Kolmos (eds.), *Engineering science, skills and bildung*, Aalborg University Press, Aalborg, 2006, pp. 43–60.
16. A. J. Bishop, International perspectives on research in mathematics education, in D.A. Grouws (ed.) *Handbook of research on mathematics teaching and learning*, Macmillan, New York, 1992, pp. 710–723.
17. H.-G. Gadamer, *Hermeneutik (Band 2): Wahrheit und Methode.—2. Ergänzungen, Register*. Mohr, Tübingen, 1993.
18. P. ten Have, *Understanding qualitative research and ethnomethodology*. SAGE Publications, London, 2004.
19. C. C. Ragin, "Casing" and the process of social inquiry, in C. C. Ragin and H. S. Becker (eds.), *What is a case?: exploring the foundations of social inquiry*, Cambridge University Press, Cambridge, 1992, pp. 217–226.
20. C. C. Ragin, *Constructing social research: the unity and diversity of method*, Sociology for a new century. Pine Forge Press, Thousand Oaks, 1994.
21. M. Borrego, Conceptual difficulties experienced by engineering faculty becoming engineering education researchers, *Journal of Engineering Education*, **96**(2), 2007, pp. 91–102.
22. M. Niss, What is quality in a PhD dissertation in mathematics education?, *Nordic Studies in Mathematics Education*, **15**(1), 2010, pp. 5–23.
23. H. Eybe and H.-J. Schmidt, Quality criteria and exemplary papers in chemistry education research, *International Journal of Science Education*, **23**(2), 2001, pp. 209–225.
24. Y. Dittrich, M. John, J. Singer and B. Tessem, For the Special issue on qualitative software engineering research, *Information and Software Technology*, **49**(6), 2007, pp. 531–539.
25. Á. Halldórsson and J. Aastrup, Quality criteria for qualitative inquiries in logistics, *European Journal of Operational Research*, **144**(2), 2003, pp. 321–332.
26. S. J. Tracy, Qualitative quality: Eight "big-tent" criteria for excellent qualitative research, *Qualitative Inquiry*, **16**(10), 2010, pp. 837–851.
27. S. Larsson, Om kvalitet i kvalitativa studier, *Nordisk Pedagogik*, **25**(1), 2005, pp. 16–35.
28. S. Larsson. On quality in qualitative studies, in *Proceedings of 1st International Workshop: Biographical research in Social and Educational Sciences*, Bremen, 1998.
29. J. Haglund, *Analogical reasoning in science education—connections to semantics and scientific modelling in thermodynamics*. Linköping University: Studies in Science and Technology Education, No. 60, Linköping, 2012.
30. J. Walther, A. L. Pawley and N. W. Sochacka, *Exploring Ethical Validation as a Key Consideration in Interpretive Research Quality*, in *ASEE annual conference*, 2015, ASEE Conferences: Seattle, Washington.
31. M. Sandelowski, A matter of taste: evaluating the quality of qualitative research, *Nursing Inquiry*, **22**(2), 2015, pp. 86–94.
32. K. R. Popper, *The logic of scientific discovery*, Hutchinson, London, 1972.
33. T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1962.
34. S. Lavelle, Technology and engineering in context: Analytical, phenomenological and pragmatic perspectives, in S. Hyldgaard Christensen, B. Delahousse, and M. Meganck (eds.), *Engineering in context*, Academica, Aarhus, 2009, pp. 75–95.
35. J. Bernhard, Engineering education research as engineering research, in S. Hyldgaard Christensen, C. Didier, A. Jamison, M. Meganck, C. Mitcham and B. Newberry (eds.), *International perspectives on engineering education: Engineering education and practice in context, volume 1*, Springer, Cham, 2015, pp. 393–414.
36. J. Hess and J. Strobel, Indigenous ways of doing: Synthesizing scholarly literature on ethno-engineering, *International Journal of Engineering, Social Justice, and Peace*, **2**(2), 2013, pp. 55–80.
37. J. Bernhard, Humans, intentionality, experience and tools for learning: Some contributions from post-cognitive theories to the use of technology in physics education, *AIP Conference Proceedings*, **951**, 2008, pp. 45–48.
38. C. S. Peirce, How to make our ideas clear, *Popular Science Monthly*, **12**, 1878, pp. 286–302.
39. J. Dewey, Experience and nature, in J.A. Boydston (ed.) *John Dewey: The later works*, Southern Illinois University Press, Carbondale, 1981, pp.
40. J. Dewey, Logic: The theory of inquiry, in J.A. Boydston (ed.) *John Dewey: The Later Works*, Southern Illinios University Press, Carbondale, 1986, pp.
41. L. A. Hickman, *John Dewey's pragmatic technology*. Indiana University Press, Bloomington, 1990.
42. L. A. Hickman, *Philosophical Tools for Technological Culture: Putting Pragmatism to Work*, Indiana University Press, Bloomington, 2001.
43. D. Ihde, *Postphenomenology and technoscience: The Peking university lectures*. State University of New York Press, Albany, 2009.
44. F. Marton, Phenomenography: Describing conceptions of the world around us, *Instructional Science*, **10**, 1981, pp. 177–200.
45. F. Marton, *Necessary conditions of learning*, Routledge, New York, 2015.
46. H. Garfinkel, *Studies in ethnomethodology*, Prentice Hall, New York, 1967.
47. L. Langsdorf, Why postphenomenology needs a metaphysics, in R. Rosenberger and P.-P. Verbeek (eds.), *Postphenomenological investigations: Essays on human—technology relations*, Lexington books, Lanham, 2015, pp. 45–54.
48. R. Rorty, *Objectivity, relativism, and truth: Philosophical papers volume 1*. Cambridge University Press, Cambridge, 1991.
49. D. Silverman, *Interpreting qualitative data: Methods for analysing talk, text and interaction*. 2nd ed., SAGE, London, 2001.
50. A. Peshkin, The goodness of qualitative research, *Educational Researcher*, **22**(2), 1993, pp. 23–29.
51. Vetenskapsrådet, *God forskningssed [Good research practice]*, 2011, Vetenskapsrådet: Stockholm.
52. D. N. Boote and P. Beile, Scholars before researchers: On the centrality of the dissertation literature review in research preparation, *Educational Researcher*, **34**(6), 2005, pp. 3–15.
53. K. Howe and M. Eisenhart, Standards for qualitative (and quantitative) research: A prolegomenon, *Educational Researcher*, **19**(4), 1990, pp. 2–9.
54. K. L. Henwood and N. F. Pidgeon, Qualitative research and psychological theorizing, *British Journal of Psychology*, **83**(1), 1992, pp. 97–111.
55. J. W. Creswell, *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*, 4th ed., Pearson, Boston, MA, 2012.
56. F. Erickson and K. Gutierrez, Comment: Culture, rigor, and science in educational research, *Educational Researcher*, **31**(8), 2002, pp. 21–24.
57. Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*, SAGE Publisher, Newbury Park, CA, 1985.
58. A. P. Bochner, Criteria Against Ourselves, *Qualitative Inquiry*, **6**(2), 2000, pp. 266–272.
59. T. A. Schwandt, Farewell to criteriology, *Qualitative Inquiry*, **2**(1), 1996, pp. 58–72.
60. J. W. Creswell, *Qualitative inquiry and research design: Choosing among five traditions*. SAGE Publications, Thousands Oaks, CA, 1998.
61. L. J. Cronbach and P. Suppes, *Research for tomorrow's schools: disciplined inquiry for education. Report on the Committee on educational research of the National academy of education*, Collier-Macmillan, London, 1969.

Jonte Bernhard is a Full Professor in Engineering Education at Linköping University, Sweden, and guest professor at KTH—Royal Institute of Technology, Sweden. He has an MSc in Engineering (Material science) and a PhD in Solid State

Physics (magnetic materials) both from Uppsala University. Before this appointment, he taught courses in engineering physics and electrical engineering for almost 25 years. He has published extensively on topics related to Engineering and Physics Education Research as well in Material Science, with more than 120 published papers and book chapters. In education research his speciality is classroom and design-based-research. He has especially focused on designing learning environments using variation theory and studying students' conceptual learning and courses of action in laboratory settings using methods inspired by interaction and conversation analysis. Presently he is associate editor for the *European Journal of Engineering Education*, chairman of SEFI's working group for engineering education research and he has been chairman of the Swedish Science Education Research Association.

Caroline Baillie is Chair of Engineering Education for the Faculty of Engineering, Computing and Mathematics at UWA and Winthrop Professor in the School of Civil, Environmental and Mining Engineering. Before coming to Perth, she held Faculty positions at Queens University, Canada, Imperial College, UK and the University of Sydney where she commenced her academic career as a materials engineering lecturer in 1992. Following a Masters in Higher Education, Baillie has, in addition to being an engineering lecturer and practitioner, been both an education developer and researcher. Her education research has focussed on threshold concepts, creative thinking, and her main passion of Engineering and Social Justice: looking at not only how students learn but *what* and *why* they learn. Baillie is a Highly Cited Researcher with a total of 23 scholarly books, and over 200 book chapters, peer reviewed journal and conference papers.