



Perspective article

How should we (not) judge the ‘quality’ of qualitative research? A re-assessment of current evaluative criteria in International Business

Catherine Welch^{a,*}, Rebecca Piekkari^b^a The University of Sydney Business School, Discipline of International Business, NSW, 2006, Australia^b Aalto University, School of Business, Department of Management Studies, Runeberginkatu 14-16, FI-00100, Helsinki, Finland

ARTICLE INFO

Keywords:

Qualitative research
Evaluative criteria
Assessment
Quality
Validity
Reviewing
Publishing

ABSTRACT

In this paper, we initiate a debate about evaluative criteria (such as validity) which are – or should be – in use to assess the quality of qualitative manuscripts in International Business (IB). We identify three generations of evaluative criteria, each derived from different philosophical orientations. Based on an analysis of published articles in two IB journals, expert interviews and sample reviews, we show how these generations shape what the scholarly community considers to be “good” qualitative research. As an alternative to rigid application of a single set of quality procedures, we advocate a pluralist, contextual approach, reflecting the inherent characteristics of IB as a field.

1. Introduction

Journals such as *The Journal of World Business* state that they welcome papers based on qualitative as well as quantitative methodologies, so long as they are “rigorous” and appropriate. But how should IB researchers decide whether a qualitative manuscript is rigorous or not? While this is a fundamental question for our research community, there is by no means agreement on the answer. The perspective we take in this paper is that settling on a single set of criteria for judging qualitative research is neither possible nor desirable. Ultimately, the best assurance of quality is a lively, reflective and open debate about the standards by which we as a research community judge what is warrantable knowledge. Our aim is to provide the ingredients for such a debate in IB by legitimizing diverse approaches to evaluating quality, and offering a way forward for them to coexist fruitfully.

In addressing the standards we use to judge the quality of qualitative research in IB, we conceptualize them as social artefacts. Evaluative criteria are produced by the research community itself: its institutional structure and traditions, and dominant intellectual influences and practices. In this regard, quality standards are as much the product of socialization, habit and convention as they are of conscious reflection and debate (Vandenberg, 2006). They are accepted as given, rather than scrutinized. These standards represent often unexamined values and beliefs about what the research community “should” do and what proper knowledge “must” look like. As such, they are prone to distortion, misapplication and misunderstanding: in short, myth-making. As Lance and Vandenberg (2015) have catalogued, it is easy for

“methodological myths and urban legends” to gain ground in a scholarly community. The only assurance we have against myth-making is to encourage an active debate about, and critical assessment of, the standards we apply (see, Sinkovics, Penz, & Ghauri, 2008).

While Vandenberg and his colleagues examine quantitative myths, we would suggest that the potential for myth-making is even greater in qualitative research, for a variety of reasons. First, its minority status in the field of IB (Piekkari, Welch, & Paavilainen, 2009) means that many reviewers are not sufficiently familiar with qualitative research to provide well-informed assessments of its quality. Second, qualitative research practices – even popular methodologies such as the case study – are less “codified” (Yin, 2014) than statistical approaches. Third, qualitative research is itself a broad church, with no single set of research standards applying to all its traditions (Easterby-Smith, Golden-Biddle, & Locke, 2008). While qualitative methodologies have in common that – to use a popular definition – they “seek to ... come to terms with the meaning, not the frequency, of ... phenomena in the social world” (Van Maanen, 1979, p. 520), they differ on the nature of “meaning” and how it should be captured. The co-existence of multiple paradigms and philosophical traditions under the broad umbrella of “qualitative research” (Prasad, 2005) has not been well understood or appreciated in IB. Yet given this diversity, there is a need for a more nuanced understanding of what constitutes “good” qualitative research and how to assess it on its merits. Imposing a single set of evaluative criteria paradoxically militates against quality being achieved: high quality and innovative manuscripts may be rejected, or serious compromises and sacrifices made in the course of the review process as

* Corresponding author.

E-mail addresses: catherine.welch@sydney.edu.au (C. Welch), rebecca.piekkari@aalto.fi (R. Piekkari).

authors attempt to squeeze their study into the dominant template despite the inconsistencies this may produce.

Since we will be advocating self-reflection and the surfacing of assumptions, beliefs and value judgements, it is therefore appropriate that we commence by being open about our own paradigmatic standpoint. We take a critical realist position, which combines a realist ontology with a subjectivist epistemology.¹ That is to say, it recognizes that our attempts at understanding the social reality external to us necessarily remain partial and imperfect. Knowledge is inherently fallibilistic: because we cannot observe the world free from our own or others' preconceptions, we can only improve the quality of our conclusions by constantly testing them against data, against rival explanations, and against the judgements of colleagues in our scientific community. Moreover, we need to acknowledge that the right practices and standards to follow are context dependent rather than universally applicable (Maxwell, 2012). This perspective helps to inform our conclusion that what constitutes best practice depends on the researcher's own paradigmatic stance, the aims of the research and the cultural practices of the research setting at that point in time.

The debate about the quality of qualitative research, to which we are contributing, is a longstanding one in the social sciences. In the remainder of the paper, we turn to this “methodological history”,² outlining three generations of evaluative criteria dating back to the 1950s, and the extent to which they have influenced IB research. Our analysis of evaluative practices in IB draws on a variety of sources: methodological texts, reviews of qualitative work solicited from IB researchers, expert interviews with qualitative IB scholars, and a close qualitative textual analysis of “validity talk” in published articles in IB journals (see Appendix A for more detail on the data sources). We argue that not only does a “third-generation”, pluralist approach to quality criteria have the potential to enrich IB, but the IB field can contribute methodological insights to this third generation. That is, context sensitivity, which IB researchers need to exercise when conducting cross-border fieldwork, is also required in validity debates; otherwise, the third generation risks perpetuating the myth of proceduralism: mistaking conformity to templates for research quality.

2. Three generations of evaluative criteria

Social scientists in the US have been seeking to formalize standards for evaluating research since at least 1950, when the American Psychological Association (APA) set up a committee to codify the proper validation of psychological tests (APA, 1954; Cronbach & Meehl, 1955). Since this first initiative, standards have been debated, disseminated to other social sciences (including management and IB), and transformed in the process. In this section, we trace this methodological history with specific reference to the evaluation of qualitative research. On the basis of our reading of the methodological literature in the social sciences, we differentiate three distinct generations of evaluative criteria. We term them “generations” rather than “stages” because while they emerged at different points in time, they can and do co-exist; although as we shall see, this often leads to inter-generational misunderstandings.

As Table 1 shows, the source of misunderstanding is paradigmatic in nature. The generations are derived from very different philosophical orientations towards comprehending the social world. Implicitly or explicitly, these orientations shape the criteria we use to judge the quality of a paper, as well as the methodological procedures that are regarded as necessary to meet these criteria. Paradigmatic commitments are also reflected in the value statements contained in a review: what a paper “should”, “needs to” or “must” do. However, given the

value-laden nature of quality criteria, we also show how each generation of criteria leads to distortions, over-simplifications, biases and unexamined assertions – that is, myths – which gain circulation. We will now consider each generation in more detail. We explain the features of each generation, as found in the methodological literature, as well as investigating how and to what extent they have been applied in the IB field. We pay particular attention to the third generation. Although we argue it has the most potential to accommodate the diverse traditions of qualitative research, we find that to date it has had little influence on IB journals.

2.1. First generation: scientism (same criteria, same procedures)

First-generation reviewers judge qualitative research against the same criteria and the same procedures as those for quantitative research (for a discussion, see Pratt, 2008). These evaluative criteria can be traced to the 1950s and the APA committee on validation, which operated during the height of the influence of logical positivism on the social sciences (Table 1; for a discussion, see Cronbach, 1989; Messick, 1987). While not amounting to adherence to the formal tenets of logical positivism (Whitley, 1984 describes it rather as “naïve” empiricism), positivistic thinking became entrenched in the social sciences, precisely the same time it was falling out of favour in philosophy (Alvesson & Skoldberg, 2009). Mainstream social scientists took for granted that the observer could accurately capture knowledge – defined as regularities of social behaviour – so long as he or she maintained objectivity and followed technical procedures that approximated those of the natural sciences (Johnson & Duberley, 2000). This view of knowledge production led to the belief that the same criteria and the same accompanying procedures should be applied to assess the quality of both quantitative and qualitative research. As qualitative research does not conform to the positivist ideal of the scientific method, nor the criteria and procedures seen to produce it, its worth came into question. The result was that qualitative research and the case study (the qualitative methodology favoured by IB) fell out of favour in this post-war period (see e.g., Platt, 1992).

The prevailing view of the case study was encapsulated by Donald T. Campbell, one of the foremost post-War psychologists to contribute to codifying the criteria and procedures for validity in experimental research. Given that the case study cannot conform to these procedures, Campbell initially judged it to be wholly lacking in scientific merit. He and a co-author branded the single case study as “of almost no scientific value” due to its “total absence of control” outside experimental settings (Campbell & Stanley, 1966, pp. 6–7). “Scientism” turns qualitative research into a poor cousin of quantitative research, because it cannot reproduce the procedures followed in the natural sciences. If not rejecting the legitimacy of qualitative research outright, first-generation beliefs can lead to the myth that qualitative research should (as far as possible) adopt quantitative procedures. The risks of scientism are, first, a higher than warranted rejection rate (i.e., papers which should be accepted, are rejected); and second, poor-quality research that mimics quantitative research may get published (i.e., papers which should be rejected, are accepted). As we shall see, Campbell (1975) was later critical of his own dismissive view.

2.1.1. First-generation evaluation in IB

First-generation beliefs surrounding the quality of qualitative research resonated in IB, which as a young discipline was striving to establish its “scientific” credentials. Yves Doz (2011, p. 585) has claimed that despite a few “pockets” such as the Nordic countries, “by and large, IB developed without much benefit from qualitative research”. Qualitative research had a marginal presence even in the early period of *JIBS*, and the percentage of qualitative papers published in the journal actually fell from the mid-1980s to mid-1990s (Piekkari et al., 2009). In a doctoral consortium held in 2004, Lorraine Eden (2004) (later a *JIBS* editor-in-chief) provided an honest assessment of the state of play. One

¹ For a recent discussion of critical realism in management, see Edwards, O'Mahoney, and Vincent (2014).

² We would like to thank one of our reviewers for proposing this term to us.

Table 1
Three generations of qualitative criteria compared.

Characteristic	First generation of evaluative criteria: scientism	Second generation of evaluative criteria: qualitative positivism	Third generation of evaluative criteria: pluralism
Time period	Formalized in the 1950s (Cronbach & Meehl, 1955; Whitley, 1984)	Popularized in management by Yin (1984) and Eisenhardt (1989)	Lincoln and Guba (1985) the first major alternative; options have proliferated since then
View of qualitative research	Same criteria, same procedures as in quantitative research: All scientific research is judged by the same criteria and needs to deploy the same procedures	Same criteria, different procedures: Qualitative research can meet the same criteria as quantitative research, but by deploying different procedures	Multiple criteria, multiple procedures: Quality criteria and appropriate procedures vary depending on the paradigm being followed
Examples of procedures	Random sampling, calculation of inter-rater reliability scores, use of software to systematize data coding	Multiple sources of evidence, use of case study database, explanation building	Procedures vary depending on the paradigmatic basis of the research
Key myth	Mimicking quantitative research: The path to better qualitative research is to conform to quantitative procedures as much as possible	Single qualitative template: Good qualitative research requires adhering to the Eisenhardt roadmap for doing case studies	Proceduralism: A quality study results from following the right procedures
Threat to quality	Applying procedures that are inappropriate to qualitative research	Restricting the diversity and innovativeness of qualitative research	Entrenching formulaic qualitative research
Examples of value statements from IB reviews	• The sample needs to be representative • The larger the number of cases, the better • In order to counteract subjectivity, software programmes need to be used	• A quality study should be exploratory • A qualitative manuscript should include propositions • Multiple case studies are superior to single case studies because they offer a better basis for making generalizations	• None found

of the top ten “don’ts” she shared with participants on getting published was doing a case study, which she described as “a “red flag” – hard to get published in most top journals”.

Given that journal reviews are not made public, it is not known how widespread this kind of scepticism towards qualitative research still is, or to what extent it contributes to the continuing low rate of qualitative work being published in *JIBS* (see Appendix A). However, we can provide evidence for the types of commentary that these beliefs can generate in the review process. In 2012³ and 2015 we solicited our colleagues’ experiences of getting qualitative research published in IB journals (see Appendix A for a description). Among these experiences were reports of reviewers rejecting a study simply because it used a qualitative approach. For example:

“This finding is based on interviews with some senior managers in five unidentified Irish subsidiaries, now in the ICT sector. The research arguments are based entirely upon unreported interviews which are not subject to scientific verification by this referee or journal.”

An extreme example was a reviewer who expressed the opinion (in contravention of the journal’s editorial policy) that the unscientific nature of the qualitative approach made the paper an inappropriate fit with an IB journal:

“As a developmental comment, I would recommend that this paper go to a journal in organisation studies where this type of convenience sampling and subjective choice of managerial replies, is apparently recognized as acceptable research methodology.”

One putative “gold standard” commonly raised in the reviews shared with us is that of random sampling. To quote one review: “The study is confined to a handful of firms in one country... How do we know that the firms are strongly representative of the population in general?” This comment reveals the reviewer is not aware that a case study generalizes to theory not to a population. Moreover, the expression “a handful of firms” points to the assumption that the larger the N, the better – even though a “strongly representative” sample is not sought in qualitative research. In a similar vein, another reviewer commented that if the sample of firms was not “representative of an

industry” it at least should be selected based “on some other measurable criterion other than being a convenience sample selected by some sort of panel of unknown origin.” The contrast the reviewer is making reveals a strong (but highly contestable) value statement: statistical sampling is without any doubt the preferable procedure because of its objective, measurable nature.

The concern about the small-N nature of qualitative research was a dominant refrain in the reviews our colleagues shared. Several of our colleagues had received comments suggesting that their study “turned out to be just a set of 35 interviews” or was based on “a fairly small number of interviews”. This overlooks that in a qualitative study, interviewees are carefully selected and typically limited only to those individuals with firsthand experience about the issue at hand. Quantity rather than quality of data is valorized in such reviews:

“a sample of 4 firms (i.e., a sample of one for each quadrant) cannot allow for any variance in your analysis. I suggest that you find 2–3 more firms in each quadrant so that you could account for some variance for the sake of some generalizability. Then you may be able to develop a strong case-based study.”

This bias towards quantity over quality of cases is, we would argue, a concerning trend as it undermines the strength of a qualitative study: its capacity for depth of insight (see also, Piekkari et al., 2009).

Qualitative data and analysis are also seen to suffer from subjectivity. “There is a concern that the cases, decisions, and quotes were cherry-picked as part of selective analysis...” Qualitative researchers are susceptible to confirmation bias: “there is a bias towards always having an explanation or set of explanations that reflect the bias of the researchers in the first instance”. These flaws of qualitative data analysis are deepened by the fact that “there is no independent verification or contestation outside the set of researchers doing the work”. To combat subjectivity, some reviewers recommended authors use computer-assisted qualitative data analysis software (CAQDAS) to undertake analysis. This was seen to render the study “more rigorous and systematic”. Reviewers in this vein recommended that authors avoid using descriptive and anecdotal evidence, such as quoting directly from interviews. Our colleagues also reported being asked to calculate inter-rater reliability scores, in an attempt to demonstrate the removal of any biases (which, as we shall see, many qualitative researchers would argue is neither possible nor desirable).

Thus, qualitative researchers did not receive useful guidance from

³ We would like to acknowledge Sylvie Chetty’s involvement in the 2012 initiative.

the first generation on how to conduct “good” qualitative research.⁴ By the 1970s, however, scientism had begun to lose credibility in methodological circles in the social sciences. It was increasingly accepted that if qualitative research mimics quantitative research as closely as possible, the outcome is a poor quality study. Since this period there has been renewed interest in qualitative research, resulting in the emergence of a second generation of evaluative criteria. This generation draws on the same positivistic criteria as the first generation but proposes different procedures to meet them – ones that are tailored to qualitative research (see Table 1).

2.2. Second generation: qualitative positivism (same criteria, different procedures)

The proponents of second-generation criteria embrace traditional positivistic quality standards, but argue that qualitative research uses different procedures to meet them (see Table 1). Prasad's (2005, p. 4) term “qualitative positivism” encapsulates the essence of this generation. To illustrate, let us return to Campbell (1975, p. 182), who in the 1970s retracted his earlier condemnation of the intensive case study. In a discussion about the use of case studies in comparative research, he acknowledged he had been unaware that

“almost invariably the social scientist undertaking an intensive case study ... ends up finding out that his prior beliefs and theories were wrong... If so, it shows that the intensive cross-cultural case study has a discipline and a capacity to reject theories which are neglected in my caricature of the method.”

Campbell's quote contains an insight that was to become central to the rehabilitation of the case study: while it might not adhere to the same procedures as experimental science, the case study provides its own arsenal of validity checks.

Campbell wrote the foreword to the first edition of Yin's (1984) seminal book on the case study, and his influence is clear throughout. Yin was trained as an experimental psychologist and in his book, he subscribes to Campbell's standards of validity. Given this commitment, his approach to validity encapsulates qualitative positivism (see Table 1): that is, he subscribes to the ideal of the neutral, objective researcher whose rigorous techniques allow for convergence on the facts (Yin, 1984). Also like Campbell, he argued forcefully that the case study is able to meet these standards – but by using a different set of techniques to those employed in quantitative (particularly survey) research. He then went a step beyond Campbell to provide a catalogue of the procedures that a case researcher could use to ensure validity and reliability (for a useful discussion, see Gibbert & Ruigrok, 2010). Yin (1984) also distanced himself from the first generation by warning against inappropriately applying quantitative procedures (such as statistical generalisation) to a case study.

Yin's criteria and procedures, as well as the positivistic axioms underlying them, are provided in Table 2, along with the example of an IB study which applies them. We have selected Bruton, Khavul, and Chavez (2011) to illustrate qualitative positivism at work. In this study, the paradigmatic influences on validity claims, research design and objectives are clear. The authors appeal, both implicitly and explicitly, to second-generation quality criteria and assumptions. The research design – the incorporation of multiple cases – and the research objective – identifying predictors of high performance – are in line with the study's goal of achieving a truthful accurate and potentially generalizable snapshot of reality by means of systematic data recording, cross-checking and analysis.

⁴ For subsequent generations of quality criteria, we provide illustrative examples of papers which apply these criteria. Because so many first-generation demands are inappropriate in the context of qualitative research, we have not done so in the case of this generation.

Yin's criteria were imported into management by Eisenhardt (1989). Eisenhardt (1989, p. 546) clearly states that hers is a positivist roadmap; that is, the goal of research is “the development of testable hypotheses and theory which are generalizable across settings”. She therefore follows Yin's argument: namely, that the case study can meet the same evaluative criteria, just through different means. Eisenhardt even goes one step further: not only can qualitative research meet the standards mandated by experimental design, but in some respects it can achieve greater validity than traditional hypothesis-testing research. Like Campbell (1975), she points to the ability of qualitative fieldwork to challenge existing preconceptions, and to reject spurious relationships through an in-depth understanding of how and why constructs are connected.

Eisenhardt's (1989) roadmap was an important step in legitimizing qualitative research in management. It provided the field a transparent set of steps to follow, as well as a defence for the contribution that the case study could make. A large part of Eisenhardt's appeal was that she was able to provide a justification for qualitative research that was able to speak to, and assuage, the quality concerns of the mainstream, while at the same time allowing for various strengths of the case study to be utilized (for a lengthier discussion, see Piekkari & Welch, 2011). The “universally recognized”, “cardinal” concepts of validity and reliability (Lee, 1999, p. 146) remained unchallenged. As we shall see in the next section, Eisenhardt's roadmap has, for some reviewers, been mythologized as a rigid template that must be followed to ensure quality research – even though Eisenhardt (1989) herself acknowledges the existence of other qualitative traditions.

2.2.1. Second-generation evaluation in IB

The main finding from our review of the two IB journals is that even in the period which we analyzed – 2006–2014 – second-generation criteria and procedures dominated the discussion of quality. Positivistic criteria were invoked, along with procedures popularized by Eisenhardt and Yin. Theoretical sampling, triangulation in its different forms (particularly triangulation of data sources and investigators) and the use of multiple cases were the most commonly mentioned procedures in the papers of our dataset. Yin's specific analytical procedures for ensuring internal validity – notably pattern matching and explanation building – were sometimes employed, but without any details being provided on how these analytical steps were undertaken. More common was the practice of providing descriptions of the coding steps or including a data display of the coding structure that was arrived at.

As mentioned earlier, while we cannot accurately estimate the prevalence of specific evaluative criteria in the IB community, the sample reviews we received reveal that at least some reviewers invoke (explicitly or implicitly) Eisenhardt (1989) as their single yardstick for evaluating quality. For these reviewers, the reliance on the Eisenhardt template provided evidence of a legitimate, rigorous and systematic approach:

“This paper is lacking methodological rigor. It neither follows accepted “grounded theory” methods nor does it meet the rigor necessary of case research. The author may want to read Kathleen Eisenhardt's and R. K. Yin's work. The author makes the point that culture studies may be appropriately conducted with qualitative methods. This is certainly true but quality methods require rigorous methodological application as do quantitative methods.”

These value statements about quality made by an anonymous reviewer can be traced directly to Eisenhardt (1989). The first is to regard the case study as suited to exploratory theory-building rather than theory-testing research. As one reviewer wrote: “The study relies on a few case studies from one country. While the case study approach is laudable, *particularly for exploratory research*, it does pose major limitations [our emphasis].” The second, related, point is to insist on the inclusion of propositions, to display potential generalizability of the findings. The following comment was typical: “To improve the paper,

Table 2
Qualitative positivism.

Qualitative positivism (Yin, 1984) Core axioms • there is a single social reality that can be discovered, predicted and controlled; • observation can be value free and independent from the observed; • knowledge claims consist of time- and context-free generalisations and linear causation		Example: Bruton, Khavul, and Chavez (2011) The article is a multiple case study of 15 micro-loan recipients in two Latin American countries (Guatemala and the Dominican Republic). The aim is to explore why some borrowers achieve high performance while others fail. The main data source consists of in-depth interviews with 15 borrowers. They also confirmed factual data from the interviews with the relevant loan officers. The core axioms underlying the paper: - Generalizability is clearly the ultimate goal, e.g. ◦ when justifying the sampling of two countries, the authors state: "Together the two countries represent diverse aspects of Latin America, yet they still have a level of commonality that makes them suitable for generalizations about microlending... Although there are differences between the countries, where microlending is concerned the two nations share many similarities that allow control over extraneous variables, and at the same time ensure that results we find are not driven by the unique institutional setting of a single nation." (pp. 723–724) ◦ the paper's theoretical contribution culminates in six testable propositions "that lay the groundwork for future research" (p. 727). - Capturing, analyzing and presenting the interviewees' experiences are regarded as unproblematic, so long as correct procedures are followed to assure accuracy. Cultural, linguistic and social differences are not seen as being impediments to the collection and analysis of interviewee data. - The multiple cases and sources of evidence converge on a single explanation
Criteria Construct validity: ensuring consistency between the construct and its measurement. Internal validity: Ensuring a relationship between cause and effect has been established. External validity: specifying the domain to which the findings can be generalized. Reliability: demonstrating that the findings will hold if the study were to be replicated.		Validity criteria and procedures used in the paper: As is often the case in journal articles, validity criteria are not explicitly mentioned in the paper. However, the procedures closely match those advocated by second-generation methodologists: - The history of the loan was verified with the loan officers concerned (i.e. use of multiple sources of evidence) - In the reporting of the findings, the chain of evidence (i.e. linkages between data and conclusions) is made clear to the reader - Use of replication logic due to inclusion of 15 cases - Establishment of a case study database consisting of field notes and interviews, which were "professionally transcribed, translated, and coded" (p. 726) - Pattern matching between theoretical categories and evidence: "We built and modified the categories generated from Step 1 [i.e., identification of a priori categories in the literature] as we discovered information that was either in support of or counter to expectations. In addition, we created new categories to reflect insights that did not fit into existing categories." (p. 726)
Procedures: Construct validity: • Multiple sources of evidence • Establishing a chain of evidence Internal validity: • Pattern-matching • Explanation building • Elucidating the dynamics of relationships • Matching against conflicting literature External validity: • Replication logic (in multiple case studies) • Comparison with existing literature Reliability: use of case study protocol/database		

you might consider including testable propositions in your paper that derive from your overall model." The third value statement is to follow Eisenhardt's preference for multiple over single case studies. Her argument that multiple case studies (preferably from 4 to 10) provide a better basis for generalization is a recurring theme in reviewer statements, such as the following: "A retrospective account of a single case study is a poor methodological choice for contributing to extant research". Anxiety about the lack of external validity and generalizability seems to underlie many of the comments – despite the fact that, as Gibbert and Ruigrok (2010) argue, internal validity is more critical to the quality of a case study.

Given that qualitative research has struggled to gain legitimacy in IB, the use of Eisenhardt (1989) as a template was – and remains – attractive. While Eisenhardt's criteria have contributed to an improvement of the status of case study research and qualitative research in IB, her clarification that her approach is positivist can easily be forgotten. The danger then becomes that positivistic assumptions are imposed on all qualitative research, including those from research traditions that are alternatives to positivism. At least for some reviewers, her criteria have become the single yardstick against which all qualitative research is judged, whether it is appropriate to the particular tradition to which the researcher adheres or not. Insistence on following her roadmap, to the exclusion of all other options, means that the criteria used to assess and defend qualitative research in IB are frozen in time – more precisely, in the 1980s, when these criteria were proposed first by Yin and then Eisenhardt.

Yet in other areas of the social sciences there has been vigorous debate about the quality criteria proposed by second-generation authorities such as Eisenhardt and Yin. As the next section shows, the trend in qualitative research methodology in other social sciences since the 1980s has been to abandon the myth of a single template for doing qualitative research, in favour of recognizing the paradigmatic diversity of qualitative research traditions (see Table 1). This moves the debate beyond differences between *qualitative* and *quantitative* research, to the

multitude of traditions *within* the qualitative realm, providing an opportunity for the qualitative community of IB researchers to initiate an active discussion.

2.3. Third generation: pluralist criteriology (multiple criteria, multiple procedures)

By the early 1980s, when IB researchers were still struggling to carve out a legitimate role for qualitative research in the field, researchers in other social science disciplines were beginning to critique the assumptions behind "qualitative positivism" (Table 1). There was a growing acceptance that the plurality of qualitative research traditions rules out any convergence on a single set of evaluative criteria (see e.g., Easterby-Smith et al., 2008). Not only did this mean rediscovering intellectual traditions and paradigms beyond those of grounded theory and experimental design (such as social constructionism, phenomenology and symbolic interactionism; for a summary, see, Prasad, 2005), but it also sparked a reappraisal of the evaluative criteria that (via Yin and Eisenhardt) had been imported from the experimental laboratory. Pluralism became a concern for qualitative rather than quantitative researchers because, unlike the latter, the former are less likely to converge on a single set of paradigmatic beliefs. In this section, we review the evolution of this pluralist agenda.

Among the earliest, and most influential, attempts to produce alternative standards were those proposed by Lincoln and Guba (1985) (Table 1). Lincoln and Guba (1985, p. 294) recognized that a shift from positivist to interpretive⁵ (or, to use their term, "naturalist") paradigms necessitated different evaluative criteria: "different strokes for different folks". Different beliefs or axioms lead to different knowledge claims

⁵ We follow Prasad (2005) by using "interpretive" as an umbrella term that encompasses distinct but related traditions: notably phenomenology, pragmatism, symbolic interactionism, hermeneutics and constructionism. These traditions assert that knowledge production is inescapably an act of human interpretation.

Table 3
Naturalist paradigm.

Naturalist paradigm (Lincoln & Guba, 1985) Core axioms • reality is socially constructed; • observation is value laden and researchers cannot be independent from the social world they study; • knowledge is time- and context-bound; • causes and effects cannot be clearly distinguished as they shape each other	Example: Yagi and Kleinberg (2011) An “interpretive ethnographic” study which aims to shed light on boundary spanning roles in a Japanese-US firm. The first author was a participant observer in the firm for a period of 7 months, supplementing her field observations with 54 in-depth narrative interviews and company documents (such as newsletters, memos and organization charts).
Criteria <i>Credibility</i>: while isomorphism between reality and findings is not possible, researchers faithfully represent the multiple constructions held by research participants <i>Transferability</i>: given variation in contexts, researchers need to show familiarity with both the “sending” and “receiving” contexts to be confident that their conclusions will hold in both. <i>Dependability</i>: replicability not achievable given changes in the researcher and researched. However, changes can be documented and explained. <i>Confirmability</i>: While objectivity is not attainable, conclusions can be shown to be internally consistent and supported by the data	The <u>core axioms</u> underlying the paper: The authors explicitly adhere to interpretivism, which “assumes human action is meaningful – that is, carried out within a system of meaning that enables us to interpret the meaning of our own and others’ actions and experience. And it assumes it is possible to grasp the subjective consciousness or intent of an actor in an objective manner” (p. 635). The contribution they offer is a theoretical “re-visioning” of the conceptualization of boundary spanning.
Procedures: <i>Credibility</i>: • Prolonged engagement • Peer debriefing • Negative case analysis • Archiving of raw data (if possible) • Member checks <i>Transferability</i>: thick description <i>Dependability</i>: inquiry audit <i>Confirmability</i>: • Inquiry audit • Triangulation • Reflexivity	<u>Validity criteria and procedures</u> used in the paper: The criteria match those of Lincoln and Guba (1985), who are also cited: - “Interpretive ethnographic accounts aim for ‘verisimilitude’ or ‘truth-likeness’ rather than provable ‘truth’. Evaluation of an account rests on whether or not the reader feels the research data are realistic, and whether or not the ethnographer has utilized data persuasively, so that representations and conclusions are believable” (p. 636) - “The goal is to demonstrate the ‘transferability’ of the research findings. That is, can they be applied fruitfully to other social settings that have some similar characteristics?” (p. 636) <u>Procedures</u> that are mentioned: - prolonged engagement in the field due to participant observation - use of multiple sources of data: participant observation, interviews and company documentation - the identity of the first author, and its impact on the study, is explicitly discussed. The first author “continually reflected on ... her “Japaneseness” (i.e., reflexivity during fieldwork). Being Japanese could be both an asset and a liability”, affecting her ability to communicate with and interpret the talk and action of research participants. It also affected her ability to build trust with Japanese and non-Japanese participants (p. 638) - “an ‘analytical auditor’, the second author of the paper..., checked for verisimilitude as research analysis progressed” (p. 639) - A follow-up 12-day site visit and telephone conversations enabled member checking - the second author also visited the research site for a total of two weeks, conducting 18 interviews, enabling her interpretations to be compared with those of the first author (i.e., investigator triangulation) - field notes were taken and interviews were transcribed, resulting in “over 800 pages of transcribed and handwritten text” (p. 638) - allowing the voices of the research subjects to be heard (p. 636), not just those of the researcher(s) (credibility through multivocality) - use of thick description, i.e. grounding the analysis in “the context” as well as the “actions, feelings and sense-making” of research participants (p. 640)

and consequently different criteria, as Table 3 shows. In their naturalist manifesto, they also specified the procedures necessary for assuring a quality naturalist study. These are related to the intimacy that researchers were able to achieve with their research participants in the course of their fieldwork. One of the ultimate tests of a quality naturalist study was whether these participants recognized their world in the researcher’s capturing of it – a process of “member checking”.

Table 3 summarizes the different “strokes” Lincoln and Guba proposed in 1985 – that is, interpretive criteria, procedures and axioms – together with an example of an IB study which applies them transparently and consistently. In this example (Yagi & Kleinberg, 2011), the authors’ claim to quality is based on their in-depth understanding of the field setting. Immersion in the field enables them to accomplish their research objective: to illuminate the “lived experience” of participants. The case study design, research objective and validity claims can therefore all be linked to the naturalistic paradigm underpinning the study. Taken together, Tables 2 and 3 show how the different paradigmatic worldviews about knowledge production underlie the choices made by the authors of the two studies, including their research questions and preferences for particular quality criteria and procedures. For example, in the ethnographic study replicability is not a meaningful criterion because the researcher is seen as the research instrument (Table 3). On the other hand, in the qualitative positivist study, producing (potentially) generalizable findings in the form of propositions is commonplace (Table 2).

Lincoln and Guba (1985) were at the forefront of what has been dubbed the “criteriology debate”.⁶ The diversity of paradigms only seemed to multiply in the 1980s, given the rise in popularity of new

trends such as postmodernism and critical theory.⁷ It soon became apparent that the two paradigms contrasted by Lincoln and Guba (i.e., positivist and naturalist) were inadequate to the task of reflecting the diverse paradigmatic positions that might underlie a qualitative study. Even Guba and Lincoln (1989) quickly renounced their own naturalist criteria as not being aligned with their commitment to a fully constructionist view of knowledge production. As they explained, “although these [naturalist] criteria have been well received, their parallelism to positivist criteria make them suspect” (Guba & Lincoln, 1994, p. 114). Instead, they proposed what they term a set of “authenticity criteria”, which they deemed to be a decisive break with positivism. Others entered the debate by proposing their own lists of criteria, with the number on offer multiplying over time (Symon & Cassell, 2012).

The proliferation of evaluative criteria has evoked a variety of responses. The first is that despite the diversity of qualitative research, it is possible to converge on some “big tent” criteria upon which everyone can agree (Eisenhardt, Graebner & Sonnenschein, 2016; Tracy, 2010). The problem is that these criteria are likely to be so broad (e.g., Tracy 2010’s stipulation that good qualitative research address a “worthy topic” and provide a “significant contribution”) that they risk not being useful. The second, postmodern stance is that the consequences of relativism “must be accepted”: that knowledge is not the quest for truth, so we cannot judge how closely we have reached it (Smith & Deemer, 2000). This position has remained a marginal one. The third response, and one gaining ground in management research, is to advocate what Johnson, Buehring, Cassell, and Symon (2006) term a “contingent criteriology”. This is a reformulation of the original insight that motivated Lincoln and Guba’s (1985) “different strokes for different folks”

⁶ The term “criteriology” initially had negative connotations (Schwandt, 1996), but is now used more neutrally.

⁷ For reviews of the many philosophical traditions which have come to influence qualitative scholars, see Duberley, Johnson and Cassell (2012); Prasad (2005).

argument: namely, that a qualitative study should be assessed according to the paradigmatic tradition within which its author is operating (see Table 1). Thus, a case study following an Eisenhardt template needs to be assessed according to different criteria than a constructionist ethnography or a narrative-based case history.

Recognition of alternative criteria has been slow in management journals. Gibbert, Ruigrok, and Wicki (2008) reviewed the case studies published in 10 top management journals between 1995 and 2000, but did not find any articles that used anything other than second-generation criteria rooted in qualitative positivism. A follow-up review has not been conducted, but the use of alternative criteria by management researchers, while now appearing in major journals, is still marginal. In his edited volume *Beyond Method*, Morgan (1983) sought to stimulate a debate about methodological assumptions and practices in the field, but the book was not widely cited and did not have the impact that it deserved. Gioia and his colleagues have had some success in promoting Lincoln and Guba's (1985) "trustworthiness" criteria (e.g., Gioia, Corley, & Hamilton, 2013; Shah & Corley, 2006) – long after these criteria had been discarded by their creators. Among the alternative criteria to have been proposed, reflexivity has perhaps gained the most circulation in management (see e.g., Haynes, 2012). As Cunliffe (2003) has noted, reflexivity has interdisciplinary origins, and as a result, there are variations in how the term is understood (for a discussion, see, Alvesson & Skoldberg, 2009). However, what the different variants of reflexivity share is a rejection of the positivist ideal of objectivity as being possible or desirable. Instead, a quality study results from the researcher's ability to reflect on how his or her field interactions, philosophical commitments and theoretical preconceptions moulded the interpretations and results.

In an important contribution to the "criteriology" discussion in management, Johnson et al. (2006) argue that recognition of the diverse traditions of qualitative research need not result in a relativist stance that "anything goes". Rather, criteriology is about applying the evaluation criteria that are appropriate to a particular work, given the author's paradigmatic positioning (see Fig. 1). This pluralist approach (which they term "contingent criteriology") calls for researchers to be critically reflexive about their own philosophical commitments and to

provide space for colleagues who have different epistemological beliefs and values. As Amis and Silk (2008) point out, acceptance of diverse criteria opens up space for non-positivist alternatives. Importantly, "embracing diversity" was also the catch cry of a 2011 *Academy of Management Journal* (AMJ) editorial on qualitative research. The editors did not want to encourage conformity to a single "AMJ style" of doing qualitative research. Instead, they advocate a balancing of "flexibility and creativity" on the one hand, and "a few broad principles" (Bansal & Corley, 2011, p. 236) on the other, which they narrow down to transparency, authenticity, deep engagement and reflexivity. Other scholars have contributed to paradigmatic diversity by proposing quality criteria for particular qualitative traditions, such as critical theory and postmodernism. While we have insufficient space to discuss them all, useful reviews of alternative criteria have been provided by Johnson et al. (2006) and Symon, Cassell, and Johnson (forthcoming). However, as we detail in the next section, we find few traces of the shift towards pluralist criteria in the pages of IB journals.

2.3.1. Third-generation evaluation in IB

Despite the modest, although promising, steps in management towards acknowledging that positivistic criteria are not universally applicable, the use of alternative criteria has been largely absent in IB: our example in Table 3, Yagi and Kleinberg (2011), is a notable exception. In our dataset, we found that Lincoln and Guba (1985) were cited in a small number of articles, but mostly without acknowledging the paradigmatic challenge they represent to existing criteria. Most commonly, authors drew on Lincoln and Guba's concept of transferability, but simply equated it with generalizability. Reflexivity was occasionally referred to as a criterion, but typically it was taken to mean the "removal of biases" – the opposite to its established usage in the methodological literature: that of acknowledging biases. There were a few notable exceptions to this neglect of alternative criteria (see Table 4 for some examples). Table 4 contrasts and compares the value statements of positivistic and interpretive positions. As Table 4 shows, the latter are based on recognition of the inherent subjectivities of observations and the fallibility of human knowledge. Based on these assumptions, the researcher's own identity is a source of understanding rather than

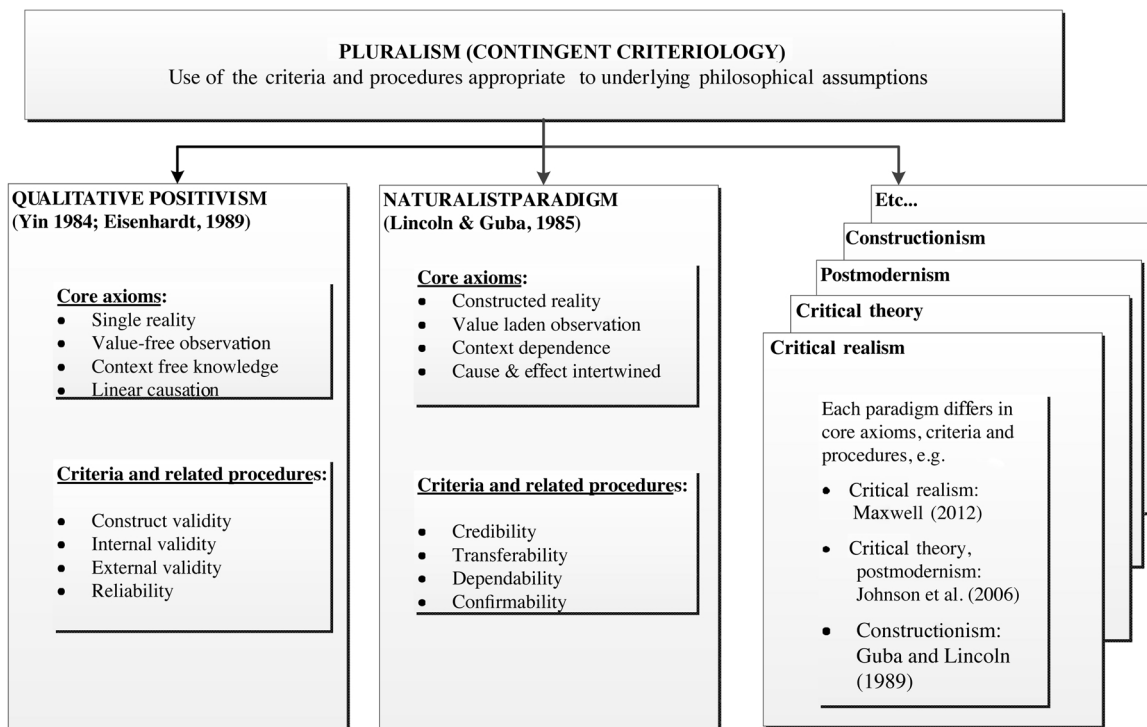


Fig. 1. Recognizing paradigmatic diversity in qualitative research.

Table 4
Positivist and interpretive value statements contrasted.

Positivist value statements	Interpretive value statements
- Personal involvement a source of bias: “Although the researcher was not an employee of the organization, the achievement of ‘objective detachment’ remained an issue, particularly in the context of regular visits.” (Pereira & Anderson, 2012, p. 225) - Triangulation to achieve single convergent account: “we triangulated the information collected from multiple sources ... [as] the basis for an objective and internally consistent description with minimal subjectivity” (Vivek, Richey, & Dalela, 2009, p. 23) - Superiority of the multiple case study: “We adopt... a multiple case study approach that generates richer theory than can a single case study” (Vallaster & Lindgreen, 2013, p. 300) - Standardisation across interviews achievable: “Reliability of measurement was ensured by consistently using one and the same interview guideline and by standardisation of the number and order of questions across all interviews” (Gassmann & Keupp, 2007, p. 354). - Standardisation across multiple coders achievable: “To further test the reliability of the coding, we compared the resultant data ..., and found no differences with regard to the overall pattern of the codes” (Barner-Rasmussen, Ehnrooth, Koveshnikov, & Mäkelä 2014, p. 893) - Accuracy of accounts assured: “Case write-ups were made available for circulation among leading participants to verify overall accuracy” (Buckley, Clegg, & Tan, 2006, p. 281)	- Personal involvement a source of interpretive insight: “We acknowledge our personal and professional interest in the subject and take our cue from ... [the] view that ‘there is no better point of entry into a critique or reflection than one’s own experience’.” (Sliwa & Johansson, 2014, pp. 1140–1141) - Triangulation to capture multiple voices: “Triangulation provided authentication, incorporated different storytellers’ perspectives, experiences and meanings and allowed multiple voices to speak” (Haley & Boje, 2014, p. 1119). - Theoretical insights provided by single case study: “The uniqueness of the case suggested it would be revelatory ..., making this case a particularly valuable research setting” (Nardon & Aten, 2008, p. 264) - Co-production not standardisation of data collection and analysis: “As researchers, we need to be aware of the active roles we play, both in the production of the narrative interviews and in the subsequent analysis of them. Narrative interviewing is not just ‘data collection’ about ‘what really happened,’ but involves a process of mutual construction of empirical material through the interviewees’ prompted storytelling about selected events. Moreover, we do not let the ‘voices of the field’ speak entirely for themselves in our analyses; we choose the quotes according to our criteria of importance from the interview transcripts.” (Gertsen & Söderberg, 2011, p. 791) - Single correct interpretation not achievable: “Along with other qualitative researchers, we do not claim that our interpretation is the only possible one, but we were careful at each stage to question our own premises for drawing the conclusions we did” (Boussebaa, Sinha, & Gabriel, 2014, p. 1158)

bias; triangulation is undertaken to capture multiple perspectives rather than settle on a single account; data and findings are co-produced by the researcher rather than representing independent facts; and single case studies are valued for their insights.

While the strength of a pluralist approach to validation is that it accommodates the diversity of qualitative research, we recognize that it nonetheless is susceptible to myth-making of its own. Although it may avoid the second-generation myth of convergence on a single positivist template, third-generation criteriology does not necessarily counter another myth to which its predecessors were also prone (Maxwell, 2012): that is, the mistaken belief that the validity of a study can be assured simply through the strict application of specific methodological procedures. Third-generation “criteriology” might even reinforce the tendency to conformity and stifle imaginative theorizing, if reviewers insist on too rigid an adherence to the set of criteria and procedures deemed appropriate to the relevant qualitative research tradition. As Eisenhardt et al. (2016) put it, the pursuit of “rigor” can descend into “rigor mortis”. In fact, the irony of any discussion on criteriology is that it risks promoting proceduralism. Therefore, we need to move beyond myth of proceduralism: how to do so is the subject of the next section.

3. Enriching the third generation: beyond the myth of proceduralism

Any exercise in establishing quality criteria is susceptible to the myth of proceduralism. Morgan (1983) cautioned against this myth as early as 1983. More recently, Alvesson and Gabriel (2013) have provided a trenchant critique of the formulaic research that proceduralism can produce. These authors argue that quality does not lie in applying the right checklist of procedures. As Vandenberg (2006, p. 200) emphasizes, this denies the inherent “grayness” and ambiguities of the research process, which cannot be reduced to a series of “if-then” procedures applied in a uniform fashion. He has labelled proceduralism as perhaps the biggest myth of them all. The danger of proceduralism, as Maxwell (2012, p. 129, quoting Phillips, 1987) points out, is that “there are no procedures that will regularly (or always) yield either sound data or true conclusions”.

In emphasizing multiple lists of criteria, third-generation criteriology may be even more susceptible to tendencies towards template-

setting. Certainly, substituting one set of procedures for another (e.g., constructionist instead of positivist guidelines) does not remove the danger of proceduralism. If poorly done, importing new criteria and procedures might even encourage crude and superficial template following, because templates may be applied without understanding their underlying assumptions and original contents.⁸ In this section, we propose context sensitivity as an important counterweight to proceduralist tendencies. Accordingly, “rigor” should not be equated with conformity to a methodological template (see also Bell, Kothiyal, & Willmott, 2017). Instead, quality results from the researcher being alert to the contextuality of the research process. Validity threats, and the most appropriate strategies for addressing them, are context dependent: that is, they not only relate to the positionality of the researcher and the objectives of the project, but also require sensitivity to the specific conditions of the research site.

Maxwell (2013), who is explicit in taking a critical realist stance, argues that the central question to address – and what ultimately is the best guide to producing a quality study – is “how might my conclusions be wrong”? Validity is ultimately about putting our interpretations to the test – a process which, at its best, results in the generation of novel theory.⁹ There are various threats that should be considered, depending on the nature of the study (Maxwell, 2012), ranging from descriptive (i.e., factual errors such as misquoting an informant), interpretive (i.e., not understanding the meaning of participants) to theoretical (e.g., not providing a clear link between evidence and claims). Thus, we suggest broadening the discussion of contingencies from the current focus on the influence of underlying philosophical axioms (e.g., Johnson et al., 2006). A truly pluralist approach to quality criteria – attuned to the challenges which IB researchers face in their fieldwork – should include the multiple ways in which decisions about validity are necessarily context dependent.

3.1. The relevance of context-dependent quality criteria to IB

The context dependence of validity decisions is relevant to any project, but we would argue it is particularly important to recognize in

⁸ We thank an anonymous reviewer for raising this point.

⁹ We would like to thank John van Maanen for encouraging us to emphasize novel and imaginative theorizing as the ultimate hallmark of a quality contribution.

IB, given the cross-border and cross-cultural nature of the domain. As is the case with any aspect of human behaviour, methodological “dos” and “don’ts” are not necessarily easily transferable across borders. The advice “it depends” applies here, too. For example, many authors in the papers we reviewed mentioned that they recorded their interviews, implying (or even explicitly stating) that this improved the accuracy of their data. However, recording may not be culturally or contextually appropriate, and may in fact reduce data quality if it means that interviewees are unable to be open in their responses. Insisting on the superiority of tape recording under these circumstances would be misguided. Researchers therefore need to be context sensitive in deciding whether methodological best practices do transfer to their field setting. They also need to apply their contextual understanding to the analysis of their data. Some authors in our dataset explicitly discuss methodological contextualization – that is, how context sensitivity was critical to producing quality data and analysis (Marschan-Piekkari & Welch, 2004). For example:

“tape recording interviews would have inhibited interviewees’ readiness to speak openly” (Gamble, 2006, p. 332, on interviewing in China)

“Topic sensitivity and social desirability impose constraints on research in China. To address these problems, the interviewer(s) deliberately asked Chinese respondents to use their group as reference rather than speaking their own mind” (Gao, Knight, Yang, & Ballantyne, 2014, pp. 314–315).

“The experience of the researcher and his or her ability to understand the phenomena studied can be an important advantage in capturing relevant data and making sense of it. My ability to speak the local language, familiarity with both American and Romanian cultures, and prior experience of working in MNCs ... greatly helped me to make sense of my observations and the participants’ narratives in a way that preserved the authenticity of their perspectives.” (Caprar, 2011, p. 612).

In the above quote, Caprar’s (2011) defence of his personal knowledge of the research context, which derived from his background and experiences, is in keeping with the ethnographic tradition in which he is operating. However, this does not mean that context dependence is only a matter for interpretivist or non-positivist traditions. In a review of published qualitative papers in (predominantly positivistic) management journals, Gibbert and Ruigrok (2010, p. 731) argue that it was in fact papers “less concerned with rigor [which] often had overly “neat” methods sections [our emphasis]”. They argue that rather than depicting a smooth, planned process, a sign of quality is for researchers to explain how they dealt with emergent issues and adapted to the context of their fieldwork. Given the highly varied contexts in which their research is conducted, IB scholars are well placed to provide considered and sensitive accounts of this process.

Acknowledging that quality criteria are context dependent does not amount to a retreat into relativism. Rather, it recognizes that research is a process of self-critique, and studies need to build in opportunities for initial conclusions and theoretical explanations to be questioned. Ultimately, a researcher needs to be able to defend why he or she has put forward particular concepts and theories as an explanation for a phenomenon. If the positivist views the researcher as a technician, the constructionist as a craftsperson, then the image of the researcher we are offering is that of a sceptic. This image acknowledges the fallibility of the human observer, and the fragility of the knowledge we produce: our conclusions may not endure for long, and it is up to the scholarly community to interrogate and improve on them. Given that validity is ultimately decided by the scholarly community, we conclude in the next section that realizing the vision of a more pluralist approach requires community-level action.

4. Conclusion: a way forward for qualitative IB research

Judging by the increasing number of qualitative articles published in *JIBS* and *JWB* in recent years (see Appendix A), it seems fair to say that qualitative research is gradually, although unevenly, penetrating the contents of the key journals in IB. However, this has occurred without a thorough discussion about how those involved – reviewers, editors, readers and authors themselves – should assess qualitative work. In this paper, we have sought to initiate such a debate. Returning to our own value judgements about knowledge production, it is our firm belief that questioning and debating the quality procedures and criteria that we take for granted are in themselves a contribution to the quality of research; they are, in fact, the mark of a thriving and successful academic community (Amis & Silk, 2008).¹⁰ In this regard, the argument we have mounted in favour of pluralism is consistent with those who have elucidated the benefits of multiparadigm inquiry (e.g., Lewis & Kelemen, 2002). Opening up a debate is particularly valuable in the IB context, we would argue, because the two major journals are largely “frozen” in the 1980s, in that the papers they publish largely do not reflect the paradigmatic diversity of qualitative research.

Disciplinary conventions, such as the prevailing evaluative criteria upheld by a field, tend to be “sticky”, meaning that changing the *status quo* requires considerable collective effort. Compared to other areas of management, institutional encouragement of qualitative research has been tentative. IB did not have an Eisenhardt who legitimized qualitative research, and its key journals did not, as has *AMJ*, provide sustained leadership in promoting understanding among readers and reviewers. The most important initiative in the two journals we reviewed for this paper was the special issue on qualitative research in IB published in *JIBS* in 2011. This milestone, which can be attributed to the leadership role played by then Editor-in-Chief, Lorraine Eden, had the explicit aim of showcasing multiple qualitative traditions (including some that have rarely appeared in IB journals, such as ethnography and narrative analysis). However, it is too early to judge to what extent this special issue produces enduring changes to how qualitative research is encouraged and evaluated (see Appendix A).

Given that in this paper we have proposed changes to existing practices in IB, it is reasonable for us to be expected to provide a pathway for achieving them. Our recommendations are not confined to journals alone: changing the evaluative criteria upheld by a field requires more than just action on the part of its journals. Criteria for judging research standards are disseminated and reproduced by a complex web of institutionalized practices: changing criteria requires changing these practices. Our recommendations might seem modest (or “tame”, to quote one of our reviewers), but we would defend them as being an achievable yet impactful agenda in a field in which diversity has to date been mostly absent. For those seeking a more ambitious action plan, Symon et al. (2017) provide a survey of reforms proposed in management.

4.1. Reforms to journal policies

To date, editors often lack detailed information about the methodological expertise that current and potential reviewers can offer. Without it, methodological matching of manuscripts becomes a mission impossible. Therefore, standard information on editorial review board members should include details of both thematic and methodological expertise, which could be explicitly used in selecting, inviting and motivating potential reviewers. Second, reviewer forms and lists of keywords should be developed so that they accommodate the diversity of research traditions. At present, some IB journals have a much more sophisticated and detailed list of key words available for quantitative rather than qualitative research. Moreover, the reviewer forms in use

¹⁰ We would like to thank Jan-Erik Vahlne for also emphasizing this insight.

are worded in such a way that the default manuscript is assumed to be a deductive, hypothesis-testing paper. Third, new governance models and organizational structures that include Associate Editors who specialize in research methods and who have equal responsibility and accountability as other members of the editorial team might further support editors in their quest for methodological pluralism. This would mean the person assigning manuscripts to editors would have sufficient methodological understanding to ensure that qualitative papers receive fair treatment.

4.2. Renewing IB education and methodological training

Recognising pluralism starts at the doctoral level by ensuring that courses cover a greater range of perspectives. Courses need to be taught in such a way that draws out the implications of philosophical commitments not just for our choice of methodological approach, but even (to return to the Bruton et al., 2011; Yagi & Kleinberg, 2011 contrast) for the type of questions that we believe are worth posing. Training should not be limited to the doctoral level, but could introduce more established scholars to new methodological perspectives and promote excellence in reviewing practices. This professional development can and, we would argue, should be done by relevant academic associations. It should not be left to individual universities, but rather is a responsibility of the IB community itself.

4.3. Actions for individual IB scholars

Our plea for pluralism implies that as individual reviewers, authors and teachers/supervisors, we should not defensively restrict ourselves – or others – to the kind of qualitative research that we or our colleagues judge as “good”. Instead, we should make space for creativity, experimentation and a dialogue between the diverse traditions of qualitative research. This means that we should not impose our own set of quality criteria on the work undertaken by others. Instead, to use a metaphor suggested by Romani, Primecz, and Topcu (2011), we need to be multilingual: that is, we should be able to assess research conducted and written in other paradigmatic languages. As reviewers, we need to become acquainted with a range of methodological traditions and philosophical debates. This requires a willingness to be open-minded and tolerant, which is not easy to achieve. At the very least, reviewers would do well to follow the advice of the experienced editor and qualitative researcher Allan Lee (1995): if you are not familiar with a particular qualitative tradition – or with qualitative research in general – it is your responsibility to be open about this in your review. He recommends that reviewers clarify in their feedback reports to the authors – not only to the editors in the section devoted to confidential comments – whether they have assessed the manuscript primarily as experts on the research topic, leaving the evaluation of the qualitative approach for others to do. Responsible reviewing makes it easier for editors, authors and other reviewers to appreciate the nature of the feedback.

To conclude, qualitative research is complex and messy. However, it should not be ignored by a field that is all about explaining phenomena

in the many contexts in which business operates. As IB scholarship is at the forefront of dealing with cross-border, cross-cultural, cross-language phenomena we should aspire to drive new methodological trends in qualitative research. In this paper, we have taken the stance that IB is extremely well placed to contribute to a more pluralist vision for “criteriology”. In particular, it has the potential to inform “third-generation” understandings of what context sensitivity means for a qualitative research project. Properly accounting for our contextuality has far-reaching implications for our views of knowledge production. It touches on perhaps the biggest myth of all: that the only warrantable scientific knowledge consists of generalizations, which we would argue is outdated. Thus, debating the evaluative criteria we use has the potential to raise many additional questions about what we mean by knowledge, and how we think we generate it. This matters not only for the community of qualitative IB scholars themselves, but above all for the well-being and vibrancy of the entire field.

Acknowledgements

We would like to acknowledge the useful comments received at the research seminars held at Uppsala University (Sweden), the University of Leeds (UK), the University of St. Gallen (Switzerland) and Stanford University (USA). We have also benefited from the feedback provided at the 13th Vaasa Conference on International Business, August 26–28, 2015 in Vaasa, Finland, ANZIBA conference, 17 February 2016 in Sydney, Australia, and the AIB Annual Meeting, June 27–30, 2016 in New Orleans, USA.

We are indebted to Pervez Ghauri, Fiona Moore and Anne-Marie Søderberg for their expert opinions on how qualitative manuscripts are evaluated in IB journals. We would also like to thank Björn Ambos, Luis Araujo, Michael Gibbert, Kathleen Eisenhardt, John Van Maanen, Lakshmi Nair and Winfried Ruigrok for their valuable comments on an earlier draft of this paper. We would like to thank the following colleagues for generously sharing with us their reviewing experiences: Eva Alfoldi, Wilhelm Barner-Rasmussen, Florian Becker-Ritterspach, Mehdi Boussebaa Pavlos Dimitratos, Christoph Dörrenbächer, Mats Ehrnrooth, Daniella Fjellström, Mike Geppert, Anne-Wil Harzing, Martin Johanson, Alexei Koveshnikov, Irina Mihailova, Dana Minbaeva, Shad Morris, Kristiina Mäkelä, Emmanuella Plakoyiannaki, Markus Pudielko, Larissa Rabbiosi, Marty Reilly, Tiina Ritvala, Ayse Saka-Helmhout, Guenter Stahl, Helene Tenzer, Esther Tippmann and Eero Vaara. We would also like to thank Jonathan Doh for his editorial guidance and our anonymous reviewers.

Finally, we are grateful to Jessica Eade for her superb skills as a research assistant and to the PhD students whom we have taught and discussed the ideas presented in this paper. Any remaining errors in the manuscript are our own doing. Rebecca Piekkari is grateful to Professor Hannele Wallenius and Professor Markku V. J. Maula for providing such an inspiring research environment at the Department of Industrial Engineering and Management, Aalto University, School of Science during her sabbatical. She would also like to acknowledge the funding provided by the Foundations of Professor Pool, Helsinki School of Economics Foundation and Marcus Wallenbergs Stiftelse in Finland.

Appendix A. Data sources

Published qualitative articles in JIBS and JWB

JIBS and JWB during the period 2006–2014 (editorials, commentaries, research notes etc were excluded) (see Table A1 for the coding results). This time period provides an update of the analysis of the qualitative papers in these two journals from 1995 to 2005 and of JIBS from 1975 to 1994 conducted by Piekkari et al. (2009). Articles were coded as conceptual or, in the case of empirical papers, as qualitative, mixed method or quantitative. We excluded from the category of “qualitative” any paper that used qualitative data as no more than an illustration, with the bulk of the paper being conceptual. We also made a decision to exclude mixed-method papers, unless the qualitative part of the paper was substantial. Typically, the quantitative part of mixed-method papers was dominant, meaning that authors did not have an incentive to address the validity of the qualitative part of the study.

Table A1

Qualitative articles published in *JIBS* and *JWB* 2006–2014.

Year	<i>JWB</i> : no. of qualitative articles	<i>JWB</i> : Total no of empirical articles	%	<i>JIBS</i> : no. of qualitative articles	<i>JIBS</i> : total no. of empirical articles	%
2006	7	23	30	1	38	3
2007	5	27	19	2	50	4
2008	9	30	30	5	59	8
2009	5	30	17	3	65	5
2010	13	30	43	3	62	5
2011	9	55	16	9	43	21 ^a
2012	9	40	23	1	29	3
2013	12	47	26	2	34	6
2014	12	44	27	10	41	24 ^a

^a All but one of the qualitative articles published in 2011 in *JIBS* appeared in a special issue on qualitative research. Half the articles published in 2014 appeared in a special issue edited by three qualitative researchers.

Reviews of qualitative papers from IB journals

In 2012, we organized panel sessions on evaluating the quality of qualitative research. As a basis for the panel discussion, we solicited reviews from about 50 qualitative researchers who had their manuscripts reviewed in IB journals. In 2015, we contacted about 40 qualitative researchers, most of whom we had not been in touch in 2012. Some of our respondents also contacted their respective colleagues whom they believed could contribute to our study. This increased the sample of qualitative researchers to 46. In total, we received 20 responses, ranging from very brief emails to extensive reflections on their reviewing experiences in IB journals. This led to a handful of additional responses. Some of the respondents also attached sample reviews and the submitted manuscript. One of these responses was so insightful that we contacted this colleague and conducted an expert interview with her over email. We complemented this with another face-to-face expert interview and a personal discussion with an Editor-in-Chief of an IB journal. Thus, in total, we had three expert interviews, 12 comments on reviewing experiences in 2012 and 20 comments in 2015.

It is important to note that we are not suggesting that the reviews we received are representative, nor are we making claims about the frequency of “methodological myths”. However, the reviews allowed us to identify the existence of a range of value judgements and beliefs about qualitative research that co-exist in the IB community, even though the confidentiality of the review process does not make it possible to judge their frequency.

References

- Alvesson, M., & Gabriel, Y. (2013). Beyond formulaic research: In praise of greater diversity in organizational research and publications. *Academy of Management Learning and Education*, 12(2), 245–263.
- Alvesson, M., & Skoldberg, K. (2009). *Reflexive methodology: New vistas for qualitative research* (2nd ed.). London: Sage.
- American Psychological Association (APA) (1954). *Technical recommendations for psychological tests and diagnostic techniques*. Washington: APA Inc.
- Amis, J. M., & Silk, M. L. (2008). The philosophy and politics of quality in qualitative organizational research. *Organizational Research Methods*, 11(3), 456–480.
- Bansal, P., & Corley, K. (2011). The coming of age for qualitative research: Embracing the diversity of qualitative methods. *Academy of Management Journal*, 54(2), 233–237.
- Barner-Rasmussen, W., Ehrnrooth, M., Koveshnikov, A., & Mäkelä, K. (2014). Cultural and language skills as resources for boundary spanning within the MNC. *Journal of International Business Studies*, 45(7), 886–905.
- Bell, E., Kothiyal, N., Willmott, H. (forthcoming). Methodology-as-technique and the meaning of rigour in globalized management research. *British Journal of Management*, doi: 10.1111/1467-8551.12205.
- Boussebaa, M., Sinha, S., & Gabriel, Y. (2014). Englishization in offshore call centers: A postcolonial perspective. *Journal of International Business Studies*, 45(9), 1152–1169.
- Bruton, G. D., Khavul, S., & Chavez, H. (2011). Microlending in emerging economies: Building a new line of inquiry from the ground up. *Journal of International Business Studies*, 42(5), 718–739.
- Buckley, P. J., Clegg, J., & Tan, H. (2006). Cultural awareness in knowledge transfer to China—The role of guanxi and mianzi. *Journal of World Business*, 41(3), 275–288.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- Campbell, D. T. (1975). Degrees of freedom and the case study. *Comparative Political Studies*, 8(2), 178–193.
- Caprar, D. V. (2011). Foreign locals: A cautionary tale on the culture of MNC local employees. *Journal of International Business Studies*, 42(5), 608–628.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147–171). Urbana: University of Illinois.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Cunliffe, A. L. (2003). Reflexive inquiry in organizational research: Questions and possibilities. *Human Relations*, 56(8), 983–1003.
- Doz, Y. (2011). Qualitative research for international business. *Journal of International Business Studies*, 42(5), 582–590.
- Duberley, J., Johnson, P., & Cassell, C. (2012). Philosophies underpinning qualitative research. In G. Symon, & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges* (pp. 16–34). London: Sage.
- Easterby-Smith, M., Golden-Biddle, K., & Locke, K. (2008). Working with pluralism: Determining quality in qualitative research. *Organizational Research Methods*, 11(3), 419–429.
- Eden, L. (2004). *Making an impact: Building a research program and getting published*. Presentation at Doctoral Student Consortium. New Orleans: Academy of Management meetings, International Management Division. <http://www.voxprof.com/eden/eden-teaching.html>.
- Edwards, P. K., O'Mahoney, J., & Vincent, S. (Eds.). (2014). *Studying organizations using critical realism*. Oxford: Oxford University Press.
- Eisenhardt, K. M., Graebner, M., & Sonenschein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal*, 59(4), 1113–1123.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.
- Gamble, J. (2006). Introducing Western-style HRM practices to China: Shopfloor perceptions in a British multinational. *Journal of World Business*, 41(4), 328–343.
- Gao, H., Knight, J. G., Yang, Z., & Ballantyne, D. (2014). Toward a gatekeeping perspective of insider-outsider relationship development in China. *Journal of World Business*, 49(3), 312–320.
- Gassmann, O., & Keupp, M. M. (2007). The competitive advantage of early and rapidly internationalising SMEs in the biotechnology industry: A knowledge-based view. *Journal of World Business*, 42(3), 350–366.
- Gertsen, M. C., & Soderberg, A. (2011). Intercultural collaboration stories: On narrative inquiry and analysis as tools for research in international business. *Journal of International Business Studies*, 42(6), 787–804.
- Gibbert, M., & Ruigrok, W. (2010). The what and how of case study rigor: Three strategies based on published work. *Organizational Research Methods*, 13(4), 710–737.
- Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, 29(13), 1465–1474.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15–31.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.
- Haley, U. C. V., & Boje, D. M. (2014). Storytelling the internationalization of the multinational enterprise. *Journal of International Business Studies*, 45(9), 1115–1132.
- Haynes, K. (2012). Reflexivity in qualitative research. In G. Symon, & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges* (pp. 72–89). London: Sage.

- Johnson, P., Buehring, A., Cassell, C., & Symon, G. (2006). Evaluating qualitative research: Towards a contingent criteriology. *International Journal of Management Reviews*, 8(3), 131–156.
- Johnson, P., & Duberley, J. (2000). *Understanding management research: An introduction to epistemology*. London: Sage.
- Lance, C. E., & Vandenberg, R. J. (Eds.). (2015). *More statistical and methodological myths and urban legends*. New York: Routledge.
- Lee, A. S. (1995). Reviewing a manuscript for publication. *Journal of Operations Management*, 13(1), 87–92.
- Lee, T. W. (1999). *Using qualitative methods in organizational research*. Thousand Oaks, CA: Sage.
- Lewis, M. W., & Kelemen, M. L. (2002). Multiparadigm inquiry: Exploring organizational pluralism and paradox. *Human Relations*, 55(2), 251–275.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Marschan-Piekkari, R., & Welch, C. (2004). *Handbook of qualitative research methods for International Business*. Cheltenham: Edward Elgar.
- Maxwell, J. A. (2012). *A realist approach for qualitative research*. Los Angeles, CA: Sage.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.
- Messick, S. (1987). *Validity. Research report*. Princeton, N.J: Educational Testing Service.
- Morgan, G. (1983). *Beyond method: Strategies for social research*. Beverly Hills, CA: Sage.
- Nardon, L., & Aten, K. (2008). Beyond a better mousetrap: A cultural analysis of the adoption of ethanol in Brazil. *Journal of World Business*, 43(3), 261–273.
- Pereira, V., & Anderson, V. (2012). A longitudinal examination of HRM in a human resources offshoring (HRO) organization operating from India. *Journal of World Business*, 47(2), 223–231.
- Piekkari, R., & Welch, C. (2011). *Rethinking the case study in international business and management research*. Cheltenham: Edward Elgar.
- Piekkari, R., Welch, C., & Paavilainen, E. (2009). The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods*, 12(3), 567–589.
- Platt, J. (1992). Case study in American methodological thought. *Current Sociology*, 40(1), 14–48.
- Prasad, P. (2005). *Crafting qualitative research: Working in the postpositivist traditions*. New York: Armonk.
- Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in valuating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods*, 11(3), 481–509.
- Romani, L., Primecz, & Topcu, K. (2011). Paradigm interplay for theory development: A methodological example with the Kulturstandard method. *Organizational Research Methods*, 14(3), 432–455.
- Schwandt, T. A. (1996). Farewell to criteriology. *Qualitative Inquiry*, 21(1), 58–72.
- Shah, S. K., & Corley, K. G. (2006). Building better theory by bridging the quantitative-qualitative divide. *Journal of Management Studies*, 43(8), 1821–1835.
- Sinkovics, R. R., Penz, E., & Ghauri, P. N. (2008). Enhancing the trustworthiness of qualitative research in international business. *Management International Review*, 48(6), 689–714.
- Sliwa, M., & Johansson, M. (2014). How non-native English-speaking staff are evaluated in linguistically diverse organizations: A sociolinguistic perspective. *Journal of International Business Studies*, 45(9), 1133–1151.
- Smith, J. K., & Deemer, D. K. (2000). The problem of criteria in the age of relativism. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 877–896). (2nd ed.). Thousand Oaks, CA: Sage.
- Symon, G., & Cassell, C. (2012). Assessing qualitative research. In G. Symon, & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges* (pp. 204–223). London: Sage.
- Symon, G., Cassell, C., Johnson, P. (forthcoming). Evaluative practices in qualitative management research: A critical review. *International Journal of Management Reviews*.
- Tracy, J. J. (2010). Qualitative quality: Eight big-tent criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851.
- Vallaster, C., & Lindgreen, A. (2013). The role of social interactions in building internal corporate brands: Implications for sustainability. *Journal of World Business*, 48(3), 297–310.
- Van Maanen, J. (1979). Reclaiming qualitative methods for organizational research: A preface. *Administrative Science Quarterly*, 24(4), 520–526.
- Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, 9(2), 194–201.
- Vivek, S. D., Richey, R. G., Jr., & Dalela, V. (2009). A longitudinal examination of partnership governance in offshoring: A moving target. *Journal of World Business*, 44(1), 16–30.
- Whitley, R. (1984). The scientific status of management research as a practically-oriented social science. *Journal of Management Studies*, 21(4), 369–390.
- Yagi, N., & Kleinberg, J. (2011). Boundary work: An interpretive ethnographic perspective on negotiating and leveraging cross-cultural identity. *Journal of International Business Studies*, 42(5), 629–653.
- Yin, R. K. (1984). *Case study research: Design and methods*. Beverly Hills: Sage.
- Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Beverly Hills: Sage.