

English-medium Programmes in Japan : A Critical Analysis of Research into Teacher Beliefs

著者	Birtles Mark
出版者	Faculty of Global and Interdisciplinary Studies, Hosei University
journal or publication title	GIS journal : the Hosei journal of global and interdisciplinary studies
volume	8
page range	163-191
year	2023-03
URL	http://doi.org/10.15002/00026274

English-medium Programmes in Japan: A Critical Analysis of Research into Teacher Beliefs

Mark Birtles

Faculty of Global and Interdisciplinary Studies, Hosei University

Abstract

This paper critically evaluates Annette Bradford's 2019 paper "It's not all about English! The problem of language foregrounding in English-medium programmes in Japan", an example from a modest body of literature concerning teacher beliefs about policy-level English-medium instruction issues in Japan. Bradford's research engages with the debate around language proficiency in these programmes and advocates for a reduced emphasis, in favour of a pedagogical and intercultural skills focus. The critique here systematically evaluates Bradford's research design using an analysis schema adapted from the four "elements" of research design outlined by Crotty (1998). The limitations of Bradford's study are discussed, with particular reference to Guba and Lincoln's (1985) trustworthiness criteria, researcher positionally and the ramifications of data reuse. This critical analysis finds that whilst Bradford's overarching conclusions and recommendations are a meaningful contribution to the literature, there are problems with aspects of the research design, ones which future scholars should take note of and make a conscious effort to avoid.

Keywords: English-medium Instruction, Japanese higher education, philosophy of research

Article Choice

Having worked for nearly a decade in an EMI environment in Tokyo, I am all too aware of how the autochthonous concept of *enryo* is ingrained in every aspect of life, complicating the educational research conducted in Japan. *Enryo* translates as a “thoughtful consideration in the literal sense of the two characters with which it is written — en, distant, and ryo, consideration” (Doi, 2014, p. 38); it symbolises the Japanese culture of restraint with words or actions aiming to maintain a harmonious atmosphere, which manifests in all kinds of ways — be it a salaryman refraining from taking a phone call on a train, or Japanese sports fans meticulously collecting and disposing of litter in a stadium after a game. From my vantage point, I can clearly see how *enryo* impacts qualitative research into the faculty side of Japanese EMI programmes: concealing personal thoughts and sentiments to maintain a status quo is *de rigueur* in Japanese universities and one of the principal reasons why in-depth studies with faculty staff about policy-level EMI issues are so rare.

This critical analysis utilises Annette Bradford’s 2019 paper “It’s not all about English! The problem of language foregrounding in English-medium programmes in Japan” (herein referred to as “the 2019 study”) as a prism for appraising research into teacher beliefs in Japanese English-medium instruction (EMI) programmes. Due to the scarcity of such research, it is vital to critically evaluate the trustworthiness of the findings of papers such as Bradford’s 2019 study as the conclusions drawn impact policy decisions and the limited number of studies are often cited as representing a national situation in cross-cultural comparisons.

The second reason I decided to choose this paper is that Bradford reused interview transcripts rather than collecting new data, which piqued my interest. The 2019 study is a reanalysis of research carried out during her 2015 EdD doctoral thesis “Internationalization Policy at the Genba: Exploring the Implementation of Social Science English-Taught Undergraduate Degree Programs in Three Japanese Universities” (herein referred to as “the 2015 study”). The nascent method of secondary qualitative analysis is increasingly accepted in scholarly communities, thanks mainly to technological advances that have made it a viable research tool (Bishop & Kuula-Luumi, 2017). There are clear scholarly and societal advantages of this data reuse method: time, money and labour can all be reduced, as can the intrusion into participants’ lives (McGinn, 2008). However, there are a considerable number of methodological and ethical challenges to overcome and, as EMI is a relatively new form of pedagogy, there must be careful consideration as to whether the secondary analysis of data is the best tool for examining a rapidly developing area of education — one a state of continual flux.

In appraising how data reuse methods are employed in this EMI research by Annette Bradford, the objective of this paper is not to discredit the arguments or recommendations she makes; instead, it is to offer an efficacious critical review, one which will be a valuable resource for scholars engaging with Japanese EMI issues and secondary data analysis in the future.

Introduction

Bradford's 2019 study is one of only a few investigating teacher perceptions of the English language within EMI programmes in Japan. Although the term EMI is used variously, the literature points to Macaro (2018) as the most salient definition of the term, defining it as "the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language (L1) of the majority of the population is not English" (Macaro, 2018, p.21). What makes EMI different from other forms of English teaching is that the central pedagogical aim is solely to impart academic content rather than "teach" the language itself; in other words, English is the vehicle rather than the cargo (see Macaro, Tian & Chu, 2020).

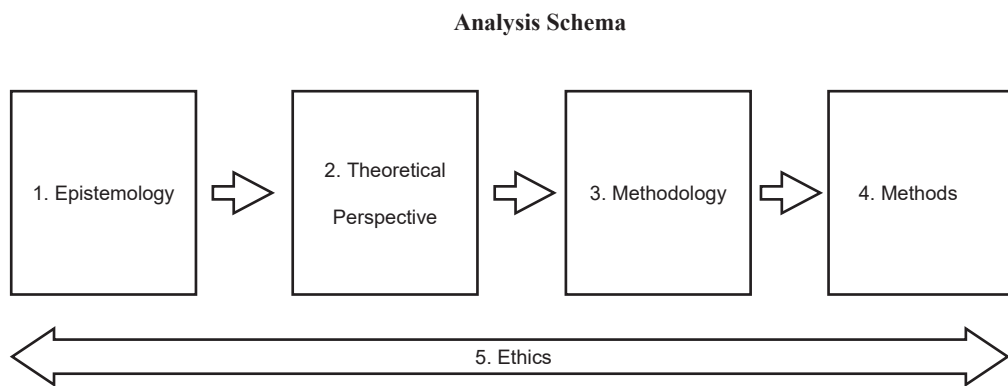
Non-Anglophone countries have seen a rapid expansion in EMI programmes in recent years and Japan is an excellent example of a typical growth pattern, driven by government targets and funding in a drive for internationalisation (Curle et al., 2020). However, Japan's unfettered EMI growth has been less than uniform, with little coordination in the roll-out and implementation of programmes, as well as a lack of consistency in the national approach to EMI (Bradford & Brown, 2018). The consensus in the literature is that there needs to be a "more integrated, comprehensive and, hence, less ad hoc approach to English-medium instruction" (Dafouz and Smit, 2020, p. 118); Bradford's 2019 study into the foregrounding of English is an attempt to make sense of some of these big-picture issues and provides a potential real-world solution.

In the study, Bradford looked at data from interviews conducted with EMI department faculty members, with her findings identifying a preoccupation with linguistic ability as one of the reasons behind the challenges faced by the programmes: English proficiency affects the content of courses, is a deterring factor in instructor recruitment and is a point of frustration for students. Bradford's paper argues that there is currently an over-emphasis on English language proficiency in Japanese EMI programmes and she recommends a shift in focus to developing pedagogical and intercultural skills within Japanese university EMI departments. This critical analysis assesses the grounds Bradford has to recommend such institutional policy changes. In

particular, this paper seeks to explore and challenge the critical philosophical underpinnings of the research process and assess to what extent the conclusions are based on reliable evidence.

Before conducting a critical analysis of Bradford’s research, attention must, for a moment, be turned to the problematic nature of terminology used in the philosophy of educational research itself. At best, specific academic vocabulary is applied less than uniformly in the literature, and at worst, it can even be at odds (Gray, 2004). Therefore, to the greatest degree possible, I have tried to make the definitions of the terminology I am applying to this critical analysis as straightforward as possible. The analysis will be broken down into sections detailed in a schema below to aid the direction and comprehension of this in-depth study of Bradford’s 2019 paper (see Figure 1). The analysis is structured to examine the four “elements” of research design outlined by Crotty (1998). First, the epistemology of the research will be considered, as this is the most “macro” level, with the foundational epistemological orientation having a knock-on effect in the later stages of research design. This continues through the theoretical perspective and methodology, as these colour the decisions made at subsequent steps in the research process. The final and most “micro” level of analysis will appraise the specific methods used to collect and analyse the data.

Figure 1.



Source: adapted from Crotty (1998) p.4

There are two important things to note about the structure of this critical analysis schema. Firstly, there is no separate section for ontology, despite it being an integral part of a researcher’s outlook. This is because, as Crotty (1998) points out, “epistemological issues and ontological issues arise

together” (p.11) — they are inexorably linked and constantly in a state of interplay. Therefore, the discussion on ontology and epistemology will be carried out *together* in the first section of the analysis. Secondly, this paper includes an additional fifth section of analysis dealing with the ethical dimension of the research. Whilst this is not included in the four elements proposed by Crotty, the researcher here considers it an important point to consider as “ethics is an ever-present concern for all researchers; it pervades every aspect of the research process” (Goodwin et al., 2003, p. 567). As ethical considerations cut across all stages of the research process, from conception to conclusion, this will be dealt with after the four central elements of research design have been examined. As already stated, every effort will be made to define and clarify each of these terms before embarking on a rigorous examination in the relevant section of the critical analysis.

Finally, it is essential to clarify the aims of following the above-outlined schema for appraisal. In investigating the Annette Bradford 2019 study “It’s not all about English! The problem of language foregrounding in English-medium programmes in Japan”, this paper intends to answer three central questions:

- Are there issues with this EMI research design?
- How do these issues impact the trustworthiness of the research regarding the broader perceptions of English in a Japanese EMI environment?
- What recommendations can be made to EMI scholars conducting a similar kind of research in the future?

In making the aims of this critical analysis, applications of concepts and terminology clear to the reader, this paper can in turn be appropriately evaluated for its academic rigour and quality of insight.

1. Epistemology

Before a research process has even begun, scholars must first consider their philosophical assumptions about the fabric of reality and how this reality (or, indeed, *realities*) can be perceived, measured and understood (Cresswell, 2007; Crotty, 1998). These primary underpinning principles are based on ontological and epistemological presuppositions. Ontology relates to “the form and nature of reality” (Guba & Lincoln, 1994, p.21) and is concerned with what can be known about this state of existence: with what certainty can a researcher be sure of

how something *really* is in the world? This question is closely related to the concept of epistemology, which is the relationship between the researcher and what is ‘real’, pertaining to “*how we know what we know*” (Crotty, 1998, p.8, italics in original). The ontological and epistemological choices a researcher makes are the foundations of the study, underpinning and informing all design choices which follow.

As aforementioned, Bradford’s 2019 study reuses data from an earlier study conducted in 2015; to understand what the foundations are made of, it is critical to return to the earlier study first to consider the ontological and epistemological orientation. The epistemological stance in the 2015 study is clearly stated as “social constructionism” (Bradford, 2015, p.29): a belief that knowledge is produced through a process of interaction and discourse within society (Berger & Lookman, 1990; Schwandt, 2000; Patton, 2002; Cresswell, 2007). Social constructionism is often used interchangeably with the term constructivism (see Charmaz, 2006), but as Bradford (2015) points out, there is an important distinction as constructivism places importance on the *individual* in the process of knowledge production, but social constructionism has an “emphasis on collective meaning-making through social interaction” (Bradford, 2015, p.106). In other words, in social constructionism, knowledge production is an interactive process *all* stakeholders are engaged in.

There is no such explicit statement of epistemological position in the 2019 study, but Bradford’s paper contains credible evidence that she remains in a social constructionist stance. Bradford states that she sees beliefs as “contextually situated” (Bradford, 2019, p.709); the fact that knowledge (in this case, the understanding of beliefs) is established in this way suggests a social constructionist epistemology as it alludes to the fact that “truth, or meaning, comes into existence in and out of our engagement with the realities in our world” (Crotty, 1998, p.8). Furthermore, Bradford (2019) talks of her “flexibility” in the interview process to “pursue emerging information” (p.711). Bradford’s acknowledgement of being actively involved and influencing the research process further supports that she still considers herself a social constructionist and is implicitly part of knowledge production.

However, a key criticism can be levelled at Bradford's social constructionist epistemology, one which is intertwined with the ontological position it is paired with, the so-called “realism/relativism debate” (Burr, 2003, p.16). Painting with broad brushstrokes, ontology can be seen as a sliding scale from realism at one end, which believes there is a single “truth”, to relativism on the other, which posits there are multiple “truths” that exist relative to the context they are created (Lincoln & Guba, 1994; Crotty, 1998; Cresswell, 2007). Bradford does not

make her ontological position clear, so one must search for clues; there is a strong suggestion that she rejects a realist ontology as she is investigating “without an a priori agenda” (Bradford, 2019, p.709) (i.e. she is not looking to prove/disprove a “truth”). Furthermore, “data was coded inductively” (Bradford, 2019, p.712): Bradford allowed the codes to emerge from the body of data as it was analysed, rather than having a preconceived theory or hypothesis, which is a strong hint at a relativist standpoint.

For qualitative researchers who shun the objective reality of realism in favour of a relativist position, this exposes an inherent problem. If Bradford is aligned to a relativist ontology, then she is by proxy admitting she does not have absolute confidence in her findings, as a relativist-social constructionist paradigm acknowledges the legitimacy of countless other forms of knowledge and findings (Hammersly, 1992). Without having a claim to authority, there is a question as to the usefulness of Bradford’s research as “relativism leads to the conclusion that nothing can ever be known for definite” (Andrews, 2012). In other words, if each and every socially-constructed “truth” is equally valid, there is little intrinsic value in investigating and uncovering just one. Therefore, advocating for faculty-level change from such an epistemological position is challenging without the claim to ‘facts’.

2. Theoretical Perspective

The theoretical perspective is “the philosophical stance lying behind a methodology” and is one that “provides a context for the process involved and a basis for its logic and criteria” (Crotty, 1998, p.66). It is the stage in research that pools the trickle-down of ontological and epistemological considerations and applies them to a robust and appropriate research methodology; in short, it is the link between theory and practice.

As already concluded in the previous section, Bradford’s 2019 study lacks some clarity on her epistemological and ontological orientation, but it has been demonstrated that the evidence points towards a continuation of her social constructionist epistemology and a broadly relativist position on the realism/relativism spectrum. A pairing of a social constructionist epistemology with a relativist ontology suggests an interpretivist theoretical perspective, one which “*looks for culturally derived and historically situated interpretations of the social life-world*” (Crotty, 1998, p.67, italics in original). As interpretivism is an anti-positivist stance, one which was born out of a rejection of the notion that there is an ultimate “truth” to be discovered (as is the belief in positivism); such a theoretical perspective is a natural continuation of Bradford’s ontological and

epistemological position. Indeed, the information we have gleaned about this standpoint indicates a particular kind of interpretivism, which is detailed in figure 2:

Figure 2.

A summary of positivist and phenomenological paradigms

	Positivist paradigm	Phenomenological paradigm
Basic beliefs	The world is external and objective The observer is independent Science is value-free	The world is socially constructed and subjective The observer is a party to what is being observed Science is driven by human interests
The researcher should	Focus on facts Locate causality between variables Formulate and test hypotheses (deductive approach)	Focus on meanings Try to understand what is happening Construct theories and models from the data (inductive approach)
Methods include	Operationalizing concepts so that they can be measured Using large samples from which to generalize to the population Quantitative methods	Using multiple methods to establish different views of a phenomenon Using small samples researched in depth or over time Qualitative methods

Source: Gray (2004) p.22 (adapted from Easterby-Smith et al., 1991)

Gray’s summary pegs Bradford’s paradigm as phenomenological. It is clear she feels the world is a social construction; she readily admits she is involved in the observations and follows an inductive approach rather than formulating (and testing) a hypothesis. This paradigm predicts that such an outlook would lead a researcher to use multiple methods to attempt to understand a particular situation and this would best be done by qualitative investigation. Burr (2003) concurs with this, stating that “the insistence of social constructionism upon the importance of the social meaning of accounts and discourses often leads logically to the use of qualitative methods as the research tools of choice” (Burr, 2003, p.17).

The epistemological issues of not knowing a “truth” aside, Bradford’s choice of a phenomenological paradigm is a sound platform for investigating the question of English foregrounding in Japanese EMI programmes. In the literature, leading academics investigating Japanese society, such as Nishihara (2013), share the core belief that knowledge is born out of subjective interaction and champion the use of phenomenology as a theoretical framework for investigation, as “Japanese society is currently facing certain multicultural situations” (Nishihara, 2013, p.21). Bradford’s investigation into the foregrounding of English is a clear example of one such multicultural situation as she is trying to make sense of a transnational relationship (*ibid.*) in a university setting. Therefore, considering the culture and the context, the theoretical perspective of the study is appropriate.

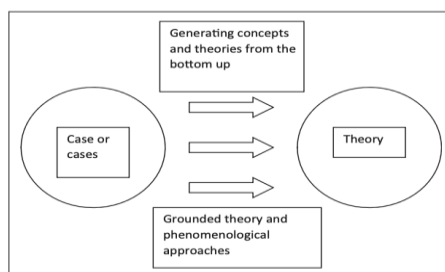
3. Methodology

Crotty (1998) defines methodology as “the strategy, plan of action, process or design lying behind the choice and use of particular methods” (p.3) and the research design Bradford chose for the 2019 study is a reexamination of a multiple case study methodology. In considering the case study as a methodology rather than a method, this critical analysis departs from Crotty’s (1998) classification for a moment, so there first needs to be an explanation as to why the case study is being treated as such. There is much discord in the literature as to whether the case study should be classified as a methodology or a method (please see Harrison et al., 2017 for an eloquent and concise discussion on the ambiguity), but here the researcher subscribes to the same view that prominent case study researchers (such as Stake, 2006; Yin 2009; Merriam & Tisdell, 2016) hold in that there is a methodology to case study research that can, in turn, employ various methods; as Harrison et al. (2017) state “This distinction accentuates the need for researchers to describe the particular underpinning methodology adopted and to clarify the alignment of chosen methods used with their philosophical assumptions and their chosen approach” (para.15).

Case study methodology has gained traction as a valid tool for investigating and comprehending multi-faceted problems and as Flyvbjerg (2006) notes, case studies are “important for the development of a nuanced view of reality” (p.223). A case study methodology with a social constructionist epistemology and a phenomenological theoretical perspective is akin to Stake (2006), who develops what Rule and John (2015) term a “theory-building approach to case study”, which can be seen in figure 3:

Figure 3.

A theory-building approach to case study



Source: Rule & John, 2015, p. 7

As this theoretical stance does not presuppose a theory, the theory is built from the ground up. Then, phenomenological methods of inquiry are employed and the theory can emerge from the data collected. Therefore, any discussion on the methodological underpinnings of the research must start with the selection of the case(s).

A researcher has a choice in case study methodology: either focus on a single case or compare multiple cases. A single case study design aims to understand a specific situation or issue, whereas a multiple case study methodology paints a more general picture (Cohen, Manion & Morrison, 2018). Bradford pursues a widespread understanding of the foregrounding of English in Japanese universities and as “one cannot generalise on the basis of a single case” (Flyvbjerg, 2006, p.224) a multiple case study seems a sensible choice for the investigation. Looking at qualitative data from several universities has the double aim of both opening up the possibility of a more nuanced interpretation and also increasing what Lincoln & Guba (1985) term “transferability”, one of their four pillars of trustworthiness. Transferability in quantitative research is an analogous term to “generalizability” in qualitative research; It is primarily a decision a reader takes, when assessing the research, of how confident they can be that results may apply to a different situation or context, but the onus is on the scholar to make every effort to produce results that can be transferred to a different setting. In Bradford's case, one can assume she selected a multiple case study methodology so that a reader can be more confident that the results are not from a single outlying case and trust that they are more broadly indicative of the situation in institutions across Japan.

However, there are two issues with the transferability in Bradford's multiple case study design. Cohen, Manion and Morrison (2018) highlight that “case studies often lack a high degree of control, and treatments are rarely controlled systematically and have little control over extraneous variables” (p.378) and as Bradford is carrying out three case studies, this issue is multiplied threefold and she offers no explanation of how she attempted to control any of the variables in the 2019 study. Furthermore, as a social constructionist, the researcher is both participant and observer in the research, which could be a possible source of bias (Shaughnessy et al., 2003). For example, Bradford's personal experience of the problem of language foregrounding could inadvertently lead the conversation in a different direction, or lead to an overstatement of the problem. Bradford offers no details of how she set about negating the potential variables in her study.

Next, Yin (2012) asserts that, as critical comparisons will be made, selecting appropriate cases is of utmost importance so that the researcher can predict similarities or differences and

develop a theory. In determining the participants for a case study, the researcher has a choice: either they choose a representative sample (by random or stratified means) or select individuals using a purposive, judgmental format (Brooks, 1997; Flyvbjerg, 2006). The findings of the 2019 study into the problem of language foregrounding draw on a larger body of data collected from the 2015 study, the main focus of which was “explor[ing] how Japanese universities are implementing undergraduate ETPs” (Bradford, 2015 p. 105). In the original 2015 study, Bradford employed a three-stage purposive sampling format: the first two stages being the selection of institutions and programmes (herein referred to as the “setting”) and the third being the selection of interviewees (herein referred to as the “participants”). In turn, the 2019 study draws on this data collected from the previous research, focussing on the interviews Bradford had conducted in 2015.

Bradford clearly states her justification for using a purposive sampling format: she uses “criterion-based selection to ensure the identification of information-rich cases” (Bradford, 2019, p.711). Purposive sampling is an excellent tactic to avoid the danger of representative sampling, which “is likely to suppress more deviant cases [...] as well as the likelihood that the full array of multiple realities [may be] uncovered” (Lincoln & Guba, 1985, p.40). Furthermore, representative sampling may also have the opposite effect in suppressing “deviant” cases and lead to the inclusion of an abundance of outlying opinions (Brooks, 1997). As well as negating these issues, purposeful sampling has the added benefit of improving the transferability of a study (Flyvbjerg, 2006) because “participants most consistent with the research design will enhance the potential that readers can assess the degree of transferability to their given context” (Jensen, 2008, p.886). However, even prominent case study researchers such as Yin (2009) acknowledge that carrying out a high-quality case study is extremely difficult, so a careful assessment of Bradford’s approach is required; this starts with an appraisal of the criterion used for selection — for both the “setting” and the “participants”.

3.1 Setting

Before entering a discussion about the selection of institutions and courses for the research, it is important to highlight two of the key terms used in the original study. In the 2015 study Bradford marks a clear distinction between EMI and what she has identified as ETPs: Bradford defines EMI as “instruction carried out through the medium of the English language” (Bradford, 2015, p.38), whereas ETPs are “HE [higher education] programs which use English exclusively as the language of instruction in countries where English is not the usual language

of instruction” (*ibid.*). EMI is positioned in the research as a catch-all term for any instruction, of any duration, given in English at the university level. In contrast, ETPs are a specific, coordinated English-only environment for the duration of an undergraduate degree.

For the 2015 study’s multiple-case study design, Bradford explicitly selected courses as three institutions offering ETPs. The first criterion was that the university in question must be participating in the Global 30 Project and have commenced an undergraduate ETP after joining the project (Bradford, 2015, p.112). This first criterion had a clearly defined purpose in the original 2015 study and cut the number of universities offering ETPs down from 27 (MEXT, 2017) to 13 candidate institutions.

However, two issues become apparent regarding the same data used in the 2019 study. Firstly, the term ETP – a term that Bradford clearly stated in 2015 was *not* analogous to the broader term EMI – is entirely absent from the 2019 study. “ETP” does not appear once in the 2019 paper; instead, the term “English-medium programmes” is used. The paper makes no indication as to whether these two terms are analogous, interchangeable or if English-medium programmes have a different definition. This in itself is problematic, but when considering the first criterion for selecting the institutions was to ensure an *explicit focus* on ETPs, the issue is drawn into sharper focus. If the 2019 paper is looking at language foregrounding in English-medium programmes, where is the justification that limiting the analysis to universities in the Global 30 Project is appropriate or relevant? In using this criterion, Bradford chooses to exclude 14 of the 27 institutions offering ETPs at the time of the 2015 study. Most notably, Waseda and Sophia universities were excluded because they had long-standing ETPs created before the inception of the Global 30 Project. However, as the research question of the 2019 study is more broadly looking at the situation in EMI, would it not have been better to consider, or even include, one of these well-established English-medium programmes? The rationale behind limiting the analysis to universities in the Global 30 project for the 2019 study is not clearly explained nor justified.

Secondly, by going further and examining the EMI situation at the time the 2019 study was published, it becomes even more apparent that there is reason to doubt that the data Bradford used was representative of institutions offering EMI in Japan at the time of the 2019 reexamination. The MEXT data for 2015 confirms Bradford’s claim that indeed 27 institutions were offering a total of 48 programmes that offered an “English-only” route to attaining a Japanese university degree (MEXT, 2017). However, when compared to the MEXT data for 2019, there had been an explosion in these kinds of programmes: at the time of the paper’s

publication, 45 institutions were offering a total of 90 English degree programmes (MEXT, 2021). This near-doubling of the number of institutions and programmes lends further credence to the argument that the first criterion used for the 2015 study meant that it was far from a representative sample in 2019. Cohen, Manion and Morrison (2018) highlight how careless sampling is “unrepresentative and unhelpful for a researcher” (p.212). Although Bradford would not have been privy to MEXT’s numbers regarding institutions providing EMI programmes (as the data for 2019 was not published until 2021), Bradford had been working at Tokyo’s Meiji University for several years at the time of conducting the 2019 study. Therefore, it is reasonable to proffer that Bradford would have been aware of the landscape at the time and should have considered the change when selecting an appropriate sample.

3.2 Participants

Next is essential to look at Bradford’s criterion for selecting individuals. The faculty participants were all teaching within EMI programmes, delivering academic content in English. Interviews with a further six staff were used to “provide context and comparative information” (Bradford, 2019, p.712). By the purposive selection of interviewees in 2015, Bradford aimed to investigate ETP from the “perspectives of senior administrators, faculty and international education support staff who have been involved in the decision-making processes and implementation of undergraduate ETPs under the Global 30 Project” (Bradford, 2015, p.115). This purposive sampling method to focus on a specific research question is a research tool that, when used judiciously, can produce illuminating data (Flyvbjerg, 2006). However, as with the selection of institutions, this creates a problem regarding the reuse of the data for the 2019 study: the focus on purposefully selecting interviewees to speak on ETPs and the Global 30 Project in the 2015 study means that Bradford was not necessarily selecting the best participants to discuss the state of EMI programmes more generally.

For the 2019 study, there is no move to try and address this issue regarding the selection of individuals for the interviews. Indeed, there is no information about the process of selecting each participant at all — this vital third criterion is simply skipped over in the discussion. This lack of detail makes it near-impossible for a reader to evaluate the quality (and, in turn, transferability) of the selection process without following the data back to the 2015 study. As already mentioned, transferability is a decision made by the reader based on the information presented in the research, but one way the researcher can aid the decision is to provide what Geertz (1973) coined a “thick description”. A thick description is about providing details of

research, but not details for detail's sake; it is to include “the cultural framework and meanings of the actors, their codes of signification, providing an emic account grounded in the actors’ cultural context” (Maxwell & Mittapalli, 2008, p.880). Bradford here is clearly culpable for not providing a thick description; important information about the selection process should not have been left out and as a result Bradford’s description lacks detail, the thickness, which allows the reader to make a better-informed decision when it comes to transferability and whether the data may fit another context.

Subsequently, consideration should be given as to whether the purposive selection process Bradford used was the best available; other studies in the literature can be used to appraise whether Bradford could have been better served by using a different selection method. Rather than using criteria to narrow down the choice and select participants, Thompson et al. (2019) used another form of purposive sampling, a maximum variation strategy, in their study of the role of self-efficacy beliefs and learner success in English medium instruction. This kind of sampling aims to collect data from the broadest possible range of perspectives and understand a problem from differing positions (Flick, 2009). The conclusions Bradford drew from her research made general recommendations about shifting emphasis from language to pedagogical and intercultural skills; using a maximal variation sampling strategy could have given a more comprehensive range of opinions on the subject.

4. Methods

Once a methodology has been established, there are an array of tools available to a researcher: these are the *methods* by which information is collected and examined in the hope of answering a research question or (dis)prove a theory (Crotty 1998). For a qualitative case study methodology, Gray (2004) suggests that questionnaires, interviews, observations and unobtrusive measures (the use of documentary evidence) are four of the most common methods used to collect primary data and indeed, the findings of Bradford’s 2019 study are based upon semi-structured interviews conducted in 2015. There is also a further choice of the method when it comes to how the data is interpreted; in Bradford’s case, this pertains to how the data was coded. Therefore, the semi-structured interview and coding methods will be appraised, with a particular focus on the specific issues related to these methods in relation to the secondary analysis of data.

4.1 Semi-structured interviews

There is consensus in the literature that qualitative case study methodologies involving semi-structured interviews have two distinct advantages: a large corpus of data can be generated with relatively little effort or input from the researcher and the method allows interviewees to control the meaning-making process, with minimal input from the interviewer (Brooks, 1997; Crotty, 1998; Scotland, 2012). As the interviewee is in a position of power concerning meaning-making, semi-structured interviews with open-ended questions are well suited to a social constructionist stance; however, some careful considerations must be taken.

First, the researcher must consider their positionality and reflexivity in an interview method as it could potentially impact the trustworthiness of the conclusions. Positionality refers to how power structures and social position can impact, and indeed shape, academic research (Lin, 2015) and reflexivity is “The monitoring by a researcher of her or his impact on the research situation being investigated” (Gray, 2004, p.405).

In Bradford’s case, the teaching staff comments about the use of English in EMI programmes could have been influenced by her positionality in the interview. Bradford is a native speaker of English and chose to conduct the 19 interviews in English (a second language for many of the interviewees) and at the time she was a professor in an EMI department at Meiji University in Tokyo. It is not difficult to see how any of these factors could have influenced the conclusion that there was a preoccupation with the foregrounding of English; the participants being interviewed in English by a native speaker who they may also be aware is a fellow EMI instructor could easily lead to a (conscious or subconscious) fixation on the role of English. Furthermore, whilst the use of open-ended questions offers the participants to provide honest and unfettered responses, semi-structured interviews in a second language can be biased by respondent articulateness (Hyman & Sierra, 2016): those who have a greater command of the interview language will be able to say more and as such, their opinion will be more present within the data when it is analysed. Those with less ability or less confidence in their English responses may perhaps have said less or articulated their opinions to a lesser degree. There is no evidence in the paper of Bradford being reflexive and questioning her positionality or assumptions regarding the data.

A final way Bradford fails to address her positionality in the 2019 study is that she makes no comment about her position as a non-Japanese interviewer. In her original study, Bradford (2015) notes that her being an outsider could either lead to open communication as she is seen as an unthreatening presence or as a barrier to communication as she lacks rapport. However, as

already outlined in the opening section of this paper, the concept of *enryo* acts as an underlying modifier to frank and open discussion in Japan, (particularly in a professional environment) (Miike, 2003), and this was most likely a defining factor in their responses — either the participants felt a release from *enryo* by speaking to a non-Japanese researcher, or their hesitation of self-expression was magnified because of it. As a quantitative researcher and social constructionist, it was vital to reflect upon this important aspect of cultural positionality as it affects the trustworthiness of the data, but Bradford fails to talk about Japanese culture, *enryo*, or its implications in either the 2015 or 2019 study.

One tactic to minimise, or ideally negate, the influence of positionality is to use “multiple methods to establish different views of phenomena” (Gray, 2004, p.22). Denzin (1970) famously termed this “triangulation”, a technique he saw as a remedy for the intrinsic biases in a single method, as “each method reveals different aspects of empirical reality” (Denzin, 1970, p.26). In Bradford’s 2019 study, there is no attempt to triangulate the conclusions she draws from the semi-structured interviews, which is problematic both for her phenomenological theoretical paradigm and her research’s trustworthiness. Referring back to figure 2., a phenomenological perspective suggests the use of triangulation to try and understand all aspects of an issue, so if Bradford is failing to triangulate her methods, then she is undermining her theoretical perspective. Alongside “transferability”, “credibility” and “confirmability” are two more of the four aspects of Lincoln and Guba’s trustworthiness criteria (1985) and are analogous to the quantitative terms “internal validity” and “objectivity”. Lincoln and Guba make clear the importance of triangulating several methods through a simple analogy:

It is as though a fisherman were to use multiple nets, each of which had a complement of holes, but placed together so that the holes in one net were covered by intact portions of other nets.

(Lincoln & Guba, 1985, p.306)

Bradford’s use of semi-structured interviews alone constitutes just one of the fisherman’s nets. Essentially, without using other methods and triangulation to cover the holes, the study may have missed important aspects of the phenomenon. Therefore, questions must be asked about the reliability of the conclusions drawn from this research.

An example of a paper that ‘covers the holes’ effectively is a study into university transitional challenges in a Japanese EMI environment conducted by Aizawa and Rose (2020).

In addition to the semi-structured interviews they conducted in a similar way to Bradford, the researchers made the crucial extra step of triangulating the findings with the addition of quantitative data: two questionnaires and two vocabulary tests. By using the additional quantitative data, Aizawa and Rose are more readily able to defend the robustness of their findings and thus better defend their credibility and confirmability. Furthermore, it is also telling that the researchers clearly state that the “interviews were conducted entirely in Japanese to ensure language did not hinder the reliability of the data” (Aizawa & Rose, 2020, p.5), which stands at odds with Bradford's decision to conduct all interviews in English.

Of course, not all researchers subscribe to the notion that triangulation is key, nor is it the only way to enhance credibility. Lincoln and Guba (1985) also suggest prolonged engagement and persistent observation as two other ways to achieve trustworthiness. In other words, in lieu of triangulation, Bradford could have chosen to interview in-depth to try and increase credibility and confirmability by scheduling multiple interviews over a period of time, which would have been another route which led to the collection of rich, thick data. Unfortunately, this represents another missed opportunity for Bradford in her research. Data continues to get richer and thicker until it reaches what Saunders et al. (2017) call the point of “data saturation”. Figure 4 details four kinds of saturation:

Figure 4.

Models of saturation

Model	Description	Principal focus
Theoretical saturation	Relates to the development of theoretical categories; related to grounded theory methodology	Sampling
Inductive thematic saturation	Relates to the emergence of new codes or themes	Analysis
A priori thematic saturation	Relates to the degree to which identified codes or themes are exemplified in the data	Sampling
Data saturation	Relates to the degree to which new data repeat what was expressed in previous data	Data collection

Source: Saunders et al., 2017, p.1897

It is difficult to see the justification that Bradford had reached the point of data saturation and data was “repeating” itself, as the conclusions in Bradford’s study were drawn off a single round of interviews with the 19 faculty members. It is reasonable to motion that in these one-time

interviews, there was still scope for additional data collection; for example, as four years had passed since the discussions had taken place, a second round of contemporary interviews would have given new, additional data which would have been pertinent for the study. Once again, Bradford here missed a chance to improve the strength of the research and aid credibility.

In continuation, the fact that so much time had passed since the data from the semi-structured interviews was being reused also raises some issues which relate back to her epistemological orientation. As Bradford states in her 2015 study, social constructionism “encourages the use of open-ended questions to allow participants to construct the meaning of phenomena” (p.122). However, when used across studies, open-ended questions present some difficulty because “choices and contexts change over time” (Hyman & Sierra, 2016, p.4). This is something that Bradford should have considered with a social constructionist epistemology, one in which “individuals construct their realities, mediated by their past experiences and cultural perspectives” (Bradford, 2015, p.137). It is clear that a social constructionist should have anticipated a change in the data considering the length of time since the original interviews. To take one example from the 2019 paper, Bradford highlights how “English limitations on the part of students and other faculty members prevented them from covering as much content as they would like” (Bradford, 2019, p.712). This is precisely the kind of place where a social constructionist should anticipate a potential shift in context and a change of feeling in the space of four years — the professors in question may now frame their contemporary experiences differently.

The reuse of data also impacts the other benefits of choosing a semi-structured interview method. One of the design strengths of semi-structured interviews with open-ended questions is that it gives the researcher flexibility to react to topics spontaneously (Patton, 2002), but this was not possible as the interviews had taken place four years previously. The collection of data on EMI would no doubt have been more fruitful if Bradford had been able to specifically probe the subject of the foregrounding of English in EMI programs, rather than the interview focus being on ETPs. Furthermore, a feature of semi-structured questions is that the prompts are designed to address the research problem (Rubin & Rubin, 2012). As the foregrounding of English was not one of the original research questions, the lines of questioning pursued during the interview were not consistent with the research question in the 2019 study.

4.2 Data Analysis

The second method that warrants comment is the secondary analysis Bradford performed in the 2019 study using the data generated from the qualitative semi-structured interviews in the 2015 study. Bradford is not alone in opting for this research method; a prolonged period of technological innovation, the movement toward open data-sharing policies and changes in how research is funded have all fuelled the move toward qualitative secondary data analysis as a quick and low-cost research method (Bishop & Kuula-Luumi, 2017). Secondary data analysis is too readily seen as something of an “easy” way to produce research, but as Bickman and Rog (2009) note, “the investigator should not assume that the level of effort to produce extant data will be small or even moderate” (p.32). Secondary data analysis is far from a straightforward procedure and here the focus will be on the coding and interpretation of the data.

In the 2019 study Bradford states the secondary analysis is “coded inductively” (Bradford, 2019, p.712), but she does not give any details of the coding process, other than it was performed using the Atlas.ti software and that the “transcripts were analytically coded to identify major themes (following Richards 2009)” (Bradford, 2019, p.711). How exactly Bradford analytically coded is a mystery, but by referring to Richards (2015) (the third edition of the text, Bradford used the second edition), it is clear that this book lacks the necessary detail to be used as a guide. Richards (2015) instructs that data should be analytically coded “considering the meanings in context, and creating categories that express new ideas about the data, coding to gather and reflect on all the data related to them” (Richards, 2014, p.112). However, there is no precise detail of how a researcher should carry this out. Indeed, as the section on analytic coding constitutes only about one page of the book, it is difficult to see how Bradford could have used Richards (2015) can be referred to as a guide.

Unfortunately, there is further evidence that there are issues with the coding process, which impacts the dependability of the findings. Dependability is the fourth and final pillar of Lincoln and Guba’s (1985) trustworthiness criteria and is analogous to the quantitative researcher’s quest for “reliability” — how sure can one be that the data is objective and neutral? From the information given in the 2019 paper, the reader must assume that Bradford conducted the data analysis alone and ignored any measures for “stepwise replication” (Lincoln & Guba, 1985, p. 317). Stepwise replication involves two or more inquiry teams conducting independent inquiries and comparing results, a process designed to enhance the objectivity, neutrality and thus the dependability of the findings.

In contrast to Bradford's study, an investigation into intergenerational redistribution in an ageing society by Prinzen (2017) is an excellent example of just how detailed the coding process can — and should — be, exemplifying a cautious approach to how data was coded for a content analysis after the interviews, and thus demonstrating the dependability of the findings. Prinzen's research clearly details steps such as using a second trained coder, discussing divergences and conducting an inter-coder reliability test as steps to scaffold the dependability of the study. On the other hand, Bradford's findings are open to criticism about potential subjectivity and bias as she worked alone; she is unable to demonstrate a comparable level of dependability in her research.

Once the data had been initially analysed, member checking is a further technique that could have been used to improve trustworthiness (Guba & Lincoln, 1985; Cresswell, 2007). Member checking involves returning to the source of the data (the participants) with the analysis results, with the explicit aim of verifying that there has been no mistake in the coding process or findings. This would have been a crucial step in the 2019 study for three key reasons. Firstly, as already stated, the data was subject to a secondary analysis, which is a tricky process in which subjectivity and bias can creep in. Secondly, this analysis was conducted four years after the interviews had taken place and there had been ample time for the participants' experiences to have altered their perspectives. Thirdly, the topic under investigation had changed from ETP to one of a more general EMI; therefore, the participants may feel their comments were not applicable or may make further qualifications.

All this is not to say that secondary data analysis in the 2019 study was the wrong choice, but the analysis needed to be more systematic to safeguard trustworthiness. One example of how to better reuse the data in the 2015 study would have been to examine the existing data with a rich and thick description of the selection criteria's possible impact on the data produced. Then, as Brooks (1997) suggests, it would be fruitful to test the robustness of the conclusion via triangulation. One method of doing this would have been to reexamine the findings from the 2015 data by carrying out another round of semi-structured interviews with a purposive selection in choosing participants the researcher feels may contradict the findings. This would not only elevate the robustness of any themes produced but may also illuminate new facets of the issue and tease out further lines of inquiry. Finally, the data should have then been coded by two independent research teams using stepwise replication to enhance the dependability of the findings.

5. Ethics

The final section of the analysis turns to the ethical dimension of the research. Discussions on ethics are far more common in qualitative research when compared to quantitative research and often centre on debates regarding the researcher's positionality and motives, as well as issues surrounding data ownership (Elliot, Holland & Thomson, 2008). The ethical considerations in qualitative research start at the planning stage and continue to the publication of the results (Brinkmann & Kvale, 2019); the ethical considerations are some of the most important for a researcher because if the position is shaky at any stage of the research, then the findings could ultimately be undermined. Here, issues pertaining to informed consent and the relationship of the 2019 study to Bradford's other research will be discussed, again with particular reference to a secondary analysis of qualitative data.

The first ethical discussion centres on informed consent in relation to the reuse of the 2015 data. In the original study, Bradford sought informed consent from the participants to use their data, something that was easy to obtain at the time of the original interview. However, Bradford does not state that the participants were consulted about the reuse of the data in 2019. Whether or not informed consent is required in secondary analysis is a source of contention in the literature, with Brinkmann and Kvale (2019) noting how informed consent is ingrained in the ethical guidelines of all research in the social science community, but Heaton (2008) drawing attention to the fact that, for secondary analysis, there are logistical and ethical difficulties to contend with when re-contacting persons who have previously taken part in a study; she notes that "people [could] have changed address or may have died; being re-contacted may also be unwelcome to some former participants" (p.512).

As the necessity for informed consent in secondary analysis is up for debate, it can be argued that the fact that Bradford is reusing her own data and the topic she is researching is closely related to that of the 2015 study lessens the apparent need to re-contact the participants. However, this does not stop a former participant from potentially disagreeing with being involved in the 2019 study. As already discussed, member checking has been suggested as a reliable way to improve the dependability of Bradford's findings, so in hindsight, it would have been prudent to recontact the participants to reaffirm their informed consent at the same time as conducting the member checking. Even if Bradford did feel that informed consent at the point of data reuse was not required, her reasoning should have been made clear in the methodology section of the 2019 paper.

Further ethical questions must also be asked about how this study relates to Bradford's other work and her data reuse. The issue here is that the 2019 study into the foregrounding of English is not the only secondary analysis Bradford has conducted with the 2015 data; there are actually two further studies engaging with the same primary data. Bradford also conducted a secondary analysis of the 2015 study in a paper published in the *Journal of Studies In International Education* entitled "Toward a Typology of Implementation Challenges Facing English-Medium Instruction in Higher Education: Evidence From Japan" (Bradford, 2016). She then reused the same primary data again in the 2017 book chapter "ROAD-MAPPING English-Medium Instruction in Japan" (Bradford & Brown 2017). Whilst it is not unusual for researchers to use sections of a doctoral thesis as the basis for paper publications, it is hard to see how one round of interviews explicitly conducted to investigate issues on ETP courses can effectively stand as a solid basis for reliable data for three different investigations, without any additional methods of data collection or triangulation.

Conclusion

The central takeaway from this critical analysis is that it underscores how intricate a qualitative investigation of this nature needs to be. Rather than invalidating Bradford's findings, the attempt here has been to call attention to some of the key issues which have presented themselves and to highlight the need for a robust research design, one which echoes the complexity of the issue at hand. The 2019 study offers a unique insight into the problem of English foregrounding in Japanese EMI programmes and the recommendation to place "less emphasis on language and greater attention to pedagogical and intercultural skills" (Bradford, 2019, p.707) is something that Japanese universities should pay careful attention to. However, elements of the research design are problematic, which calls into question the robustness of the findings.

Among the recommendations discussed in this paper, the analysis has spotlighted the need for increased trustworthiness, a careful contemplation of researcher positionality and a more complete consideration of the implications of data reuse. The findings have questioned decisions made by Bradford in selecting a setting and participants and the description she provides for the process has been demonstrated to be lacking the necessary rigour to allow other scholars to appraise adequately. Bradford also ran contrary to her phenomenological paradigm and epistemological underpinning by using interview data that was several years old: as a social constructionist must concede, all knowledge is context-dependent and conclusions drawn from interviews conducted in a 2015 EMI environment (one amid rapid expansion) have only limited

relevance in 2019. This paper has demonstrated that, at numerous points, there were more robust options available to Bradford in designing and carrying out the study.

However, despite the design problems presented, future scholars must not be dissuaded from approaching research of this level of methodological, technological and cultural complexity. Rather, this paper should serve as a forewarning to pay attention to the inherent pitfalls they must navigate when conducting research of this nature. English-language research into teacher beliefs in foreign education systems — such as Japan — is in short supply, so it is vital that these scholars aspire to the highest levels of academic veracity as gatekeepers of knowledge across linguistic, cultural and geographic boundaries.

References

- Aizawa, I. and Rose, H., 2020. High school to university transitional challenges in English medium instruction in Japan. *System* [Online], 95. Available from: <https://doi.org/10.1016/j.system.2020.102390> [Accessed 10 January 2022].
- Andrews, T., 2012. What is Social Constructionism? *Grounded Theory Review* [Online], 11(1). Available from: <http://groundedtheoryreview.com/2012/06/01/what-is-social-constructionism/> [Accessed 9 February 2022].
- Berger, P.L. and Luckmann, T., 1990. *The social construction of reality: a treatise in the sociology of knowledge*. New York: Anchor Books.
- Bickman, L. and Rog, D.J., 2009. Applied research design: a practical approach. In: L. Bickman and D.J. Rog, eds. *The SAGE handbook of applied social research methods*. Los Angeles, Calif: SAGE Publications. pp.3-43.
- Bishop, L. and Kuula-Luumi, A., 2017. Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open*. January-March 2017, pp.1-15.
- Bradford, A., 2015. *Internationalization Policy at the Genba: Exploring the Implementation of Social Science English-Taught Undergraduate Degree*. EdD Thesis. Washington D.C.: The George Washington University.
- Bradford, A., 2016. Toward a Typology of Implementation Challenges Facing English-Medium Instruction in Higher Education: Evidence From Japan. *Journal of Studies in International Education*, 20(4), pp.339-356.
- Bradford, A., 2019. It's not all about English! The problem of language foregrounding in English-medium programmes in Japan. *Journal of Multilingual and Multicultural Development*, 40(8), pp.707-720.

Bradford, A. and Brown, H., eds, 2017. *English-medium instruction in Japanese higher education: policy, challenges and outcomes*. Bristol, U.K.: Multilingual Matters.

Bridges, D., 2001. The Ethics of Outsider Research. *Journal of the Philosophy of Education*, 35(3), pp.371-386.

Brinkmann, S. and Kvale, S., 2018. *Doing interviews*. Second edition. Los Angeles; London: SAGE Publications.

Brooks, M., 1997. In-Depth Interviewing as Qualitative Investigation. In: D. Griffiee and D. Nunan, eds. *Classroom Teachers and Classroom Research. JALT Applied Materials*, [Online]. Tokyo: JALT. Available from: <https://files.eric.ed.gov/fulltext/ED415702.pdf> [Accessed 15 February 2022].

Burr, V., 2003. *Social constructionism*. Second edition. Hove, East Sussex ; New York, NY: Routledge.

Charmaz, K., 2006. *Constructing grounded theory*. London; Los Angeles, Calif: Sage Publications.

Creswell, J.W., 2007. *Qualitative inquiry & research design: choosing among five approaches*. Second edition. Los Angeles, Calif: Sage Publications.

Crotty, M., 1998. *Foundations of social research: meaning and perspective in the research process*. Oxon, U.K.: Routledge.

Curle, S., Sahan, K., Jablonkai, R., Mittelmeier, J. and Veitch, A. 2020. English in higher education - English medium Part 1: Literature review. In: N. Galloway ed. *Teaching English*. London, U.K.: British Council.

Dafouz, E. and Smit, U., 2020. *ROAD-MAPPING English medium education in the internationalised university*. London, U.K.: Palgrave Macmillan.

Denzin, N.K., 2009. *The research act: a theoretical introduction to sociological methods*. New Brunswick, N.J.: Aldine Transaction.

Dörnyei, Z., 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford; New York, N.Y: Oxford University Press.

Elliot, J., Holland, J. and Thomson, R., 2008. Longitudinal and panel studies. In: P. Alasuutari, L. Bickman and J. Brannen, eds. *The SAGE handbook of social research methods*. Los Angeles, Calif: SAGE Publications. pp.228-248.

Flick, U., 2018. *An introduction to qualitative research*. 6th edition. Los Angeles, Calif: SAGE Publications.

Flyvbjerg, B., 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12(2), pp.219-245.

Geertz, C., 1973. *The interpretation of cultures: selected essays*. New York N.Y.: Basic Books.

Goodwin, D., Pope, C., Mort, M. and Smith, A., 2003. Ethics and Ethnography: An Experiential Account. *Qualitative Health Research*, 13(4), pp.567-577.

Gray, D.E., 2004. *Doing research in the real world*. Los Angeles, Calif: SAGE Publications.

Guba, E.G. and Lincoln, Y.S., 1994. Competing Paradigms in Qualitative Research: Theories and Issues. In: S.N. Hesse-Biber and P. Leavy, eds. 2004. *Approaches to qualitative research: a reader on theory and practice*. New York, N.Y.: Oxford University Press. pp.17-38.

Hammersley, M., 1992. *What's wrong with ethnography?: methodological explorations*. Oxon, U.K.: Routledge.

Harrison, H., Birks, M., Franklin, R. and Mills, J., 2017. Case Study Research: Foundations and Methodological Orientations. *Forum Qualitative Sozialforschung* [Forum of Qualitative

Social Research] [Online], 18(1). Available from: <https://doi.org/10.17169/fqs-18.1.2655> [Accessed 15 February 2022].

Heaton, J., 2008. Secondary analysis of qualitative data. In: P. Alasuutari, L. Bickman and J. Brannen, eds. *The SAGE handbook of social research methods*. Los Angeles, Calif: SAGE Publications. pp.506-519

Hyman, M.R. and Sierra, J.J., 2016. Open- versus Close-Ended Survey Questions. *NMSU Business Outlook*, 14(2), pp.1-5.

Jensen, D., 2008. Transferability. In: L. Given, ed. *The SAGE Encyclopedia of Qualitative Research Methods*. Los Angeles, Calif: SAGE Publications. p.886

Lin, A.M.Y., 2015. Researcher Positionality. In: F.M. Hult and D.C. Johnson, eds. *Research methods in language policy and planning: a practical guide*. West Sussex, U.K.: Wiley-Blackwell. pp.21-32.

Lincoln, Y.S. and Guba, E.G., 1985. *Naturalistic inquiry*. Los Angeles, Calif: SAGE Publications.

Macaro, E., 2018. *English medium instruction content and language in policy and practice*. Oxford, U.K.: Oxford University Press.

Macaro, E., Tian, L. and Chu, L., 2020. First and second language use in English medium instruction contexts. *Language Teaching Research*. 24(3), pp.382-402.

Maxwell, J.A. and Mittapalli, K., 2008. Thick Description. In: L. Given, ed. *The SAGE Encyclopedia of Qualitative Research Methods*. Los Angeles, Calif: SAGE Publications. p.880

Merriam, S.B. and Tisdell, E.J., 2015. *Qualitative research: a guide to design and implementation*. Fourth edition. San Francisco, Calif: John Wiley & Sons.

MEXT. 2017. Daigaku ni okeru kyōiku naiyō-tō no kaikaku jōkyō ni tsuite (Heisei 27-nendo) 大学における教育内容等の改革状況について（平成27年度） [Reform status of educational content at universities (2015)]. Available from: https://www.mext.go.jp/a_menu/koutou/daigaku/04052801/1398426.htm [Accessed 11 January 2022].

MEXT. 2021. Daigaku ni okeru kyōiku naiyō-tō no kaikaku jōkyō ni tsuite (Reiwa moto nendo) 大学における教育内容等の改革状況について（令和元年度） [Reform status of educational content at universities (2019)]. Available from: https://www.mext.go.jp/a_menu/koutou/daigaku/04052801/1417336_00008.htm [Accessed 11 January 2022].

Patton, M.Q., 2015. *Qualitative research & evaluation methods: integrating theory and practice*. Fourth edition. Los Angeles, Calif: SAGE Publications.

Prinzen, K., 2017. The Moral Economy of Intergenerational Redistribution in an Ageing Society: A Qualitative Analysis of Young Adults' Beliefs in the United States. *Social Policy & Administration*, 51(7), pp.1267–1286.

Rubin, H.J. and Rubin, I., 2012. *Qualitative interviewing: the art of hearing data*. Third edition. Los Angeles, Calif: SAGE Publications.

Rule, P. and John, V.M., 2015. A Necessary Dialogue: Theory in Case Study Research. *International Journal of Qualitative Methods* [Online], 14(4), pp.1-11.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H. and Jinks, C., 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4), pp.1893–1907.

Schwandt, T.A., 2000. Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In: N.K. Denzin and Y.S. Lincoln, eds. *Handbook of qualitative research*. Los Angeles, Calif: SAGE Publications. pp.189-213.

Scotland, J., 2012. Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms. *English Language Teaching* [Online], 5(9). Available from: <https://doi.org/10.5539/elt.v5n9p9>.

Stake, R.E., 2006. *Multiple case study analysis*. New York, N.Y.: The Guilford Press.

Thompson, G., Aizawa, I., Curle, S. and Rose, H., 2019. Exploring the role of self-efficacy beliefs and learner success in English medium instruction. *International Journal of Bilingual Education and Bilingualism*, 25(1), pp.196–209.

Yin, R.K., 2009. *Case study research: design and methods*. Fourth edition. Los Angeles, Calif: SAGE Publications.

Yin, R.K., 2012. *Applications of case study research*. Third edition. Los Angeles, Calif: SAGE Publications.