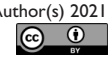
Article

Journalism
1–18


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14648849211016997
journals.sagepub.com/home/jou

Etic and emic data production methods in the study of journalistic work practices: A systematic literature review

Lauri Haapanen
University of Jyväskylä, Finland

Ville JE Manninen
University of Vaasa, Finland

Abstract

This methodological paper discusses the application of etic and emic perspectives in producing data sets for the study of journalistic praxis. The concepts refer to the researcher-analyst's and the practitioner-informant's viewpoints, respectively, and that being the case, they have inherent weaknesses if used independently. We argue that these shortcomings can be mitigated through an integrated application of the concepts and by seeing the two perspectives as complementary to each other. Combining etic and emic perspectives will allow for a more comprehensive understanding of complex yet routinized real-life behaviour such as journalistic work. Following the appraisal of the etic-emic combination methodology, we present the results of a systematic literature review. We explore the application of this methodology in the study of journalistic work, as recounted in international journal articles published between 2000 and 2017. After keyword searches into several databases and a manual review of 3018 items, a corpus of 228 relevant articles was analysed. Our results demonstrate the studies' disproportionate emphasis on the earliest phases of the journalistic process, that is, *topic discovery*, *topic selection*, *point of view selection* and *sourcing*, and the relative scarcity of studies applying a fully integrated etic-emic-combination methodology. We also report on, for example, the types of research material used and the media studied. We conclude with recommendations for expanding and diversifying the use of etic-emic-combination

Corresponding author:
Lauri Haapanen, University of Jyväskylä, PO box 35, FI-40014, Finland.
Email: lauri.m.haapanen@jyu.fi

methodology in journalism studies to give us a more balanced understanding of the field.

Keywords
Data collection, emic, etic, journalism, literature review, methodology, work practice

Introduction

All human behaviour has two components: an observable action and the intention(s) behind it. Raising one's open hand, for example, is an act that could be detected by a bystander, even by a simple motion sensor. However, interpreting it as a casual greeting, a military salute, or a request for a speaking turn requires contextual understanding on the part of the bystander – perhaps verbal explanation from the hand-lifter, too. In other words, to truly know and understand human behaviour, be it social conventions or, as in this paper, journalistic praxis, knowledge of both behaviour and its meaning are required (e.g. Olivier de Sardan, 2015: Chapter 3).

In order for academic research to support and improve journalistic practices, it is not enough for analysis to *describe* a process and/or its output. The research should also *explain* why things happen as they happen and/or why the output is as it is. In other words, research needs to present both *what* and *why* questions. The answers to these questions can be described through the concepts of *etic* and *emic* (Pike, 1954): an outsider's (etic) perspective can capture practitioners' observable behaviour but leaves researchers guessing as to its meaning, while revealing the intentions behind the behaviour requires an insider's (emic) perspective.

Both etic and emic perspectives involve epistemic shortcomings. When investigating complex and mentally demanding processes, the researcher's etic scrutiny may inadvertently neglect something relevant to the activity. Relying on emic descriptions of routine processes, on the other hand, is as if the researcher was, so to speak, outsourcing the task of generalization to the informant, who rarely possesses the necessary analytical rigour. (Haapanen, 2018; Perrin, 2013). Of course, each perspective can be sufficient for particular applications: etic for routine processes involving little discretion or conscious decision-making on the part of the informant, emic for entirely deliberate, intentional actions and one-off events. Methods must be subordinate to research goals, and therefore methods representing only one of these perspectives are effective fits for only some research questions.

The subject of this methodological study is research into journalistic work practices, and there the problems involved are quite evident. Let us illustrate it by drawing on *quoting*. To begin with, analysing ready-made quotes would not tell us much about the process of making them. Directly observing the journalist would tell us more, comparison of the quotes to the original utterances (i.e. the interview speech) would be even more revealing, and the use of computer logging software greater still. However, the material activity of quoting hardly illuminates its intention. To investigate intentions, we need to ask the journalist concerned to verbalise their thinking. An interview set-up might be

insufficient to uncover actual strategies and decision-making processes: quoting, like other routine work, is most likely conducted on auto-pilot, to alleviate the cognitive strain brought about by the hectic nature of news work. There is also the possibility that journalists will intentionally or sub-consciously dress up practices that breach guidelines or social norms (Koliska, 2015: section 8.3).

To overcome the problems described above, we propose that accurate understanding of behaviour can be achieved through the complementary perspectives of *etic* and *emic*. A research framework should carefully integrate both the researcher-analyst's and the practitioner-informant's perspectives. As regards quoting, this could be achieved by, for example, showing the journalist a transcript of an interview and the quotes they have written based on it as stimuli, and asking them to explain the recontextualization of the quoted discourse (Haapanen, 2017). The stimuli increase the journalist's awareness of their routinized performance and thus help them reconstruct their half-automatized trains of thought.

We consider a data production methodology that integrates the etic and emic perspectives to be the best way of investigating journalistic praxis, and we therefore conducted a systematic literature review of the journalism research literature to answer the following question: how often and in what ways has this kind of data production methodology been used in journalism studies between 2000 and 2017?

In what follows, we will first go through the origin and varying definitions of the concepts of *etic* and *emic* and explain our understanding of these concepts. We will then introduce the design of our literature review. The bulk of the paper is dedicated to the results. We will explain, for example, what type of media these studies investigated, what kinds of data sources they used, and what part(s) of the journalistic work process they focused on. We will also assess both the integration and balance of the etic and emic perspectives in the sample. After discussing the results, we conclude that producing diverse data sets tends to require a lot of preparation and laborious fieldwork – and is thus rarely done. However, we recommend making the effort, as it often pays off in the form of rich data with exceptional explanatory power.

Emic and *etic*: Complementary perspectives on data production

Originally, the concepts of *emic* and *etic* came from the word pair, *phonemics* and *phonetics*; the former refers to the set of sounds that are meaningful within a single language, the latter is a description of all the sounds that can be distinguished in any language. *Emic* and *etic* were coined by linguist and anthropologist Kenneth L. Pike in 1954 'as a response to the need to include nonverbal behaviour in linguistic description' (Pike, 1990: 18).

> I took the word phonemic, crossed out the *phon*-part meaning 'sound', and generalized my use of the new emic term to represent any unit of culture, at any level, of any kind, which was reacted to as a relevant unit by the native actors in that behavior. In the same way, I created the word *etic* from *phonetic*. (Pike, 1988: 154, 155)

In other words, etic refers to data that are descriptive and can reliably be produced even by someone with limited contextual understanding. Emic, in turn, refers to data that are explanatory and that refer to the meanings participants themselves attach to their experience. Emic data can be reliably produced only from a contextually grounded viewpoint.

In anthropology, the concepts of *emic* and *etic* became such well-established features of the vocabulary within the following decades that they were often no longer italicized, or necessarily sourced or defined (Headland et al., 1990: 6). At the same time, their analytical power also faded: 'Unfortunately, emic and etic have become slogans or catchwords in anthropology, rather than clear-cut concepts' (Crane and Angrosino, 1984: 125).

By the 1980s, the concepts were increasingly used in other fields, especially in psychology but also in linguistics, intercultural research, ethnography, sociology and medicine, among others (Headland et al., 1990). They were given very different definitions from one researcher to another, and from one field to another. Hahn et al. (2011: 4. See also Janićijević, 2015) list a variety of dichotomies to which emic and etic have been used to refer: specific/universal, verbal/nonverbal, interview/observation, subjective knowledge/scientific knowledge, good/bad, ideal behaviour/actual behaviour, description/theory, soft facts/hard facts – to name but a few.

As definitions vary both diachronically and synchronically, that is, over time and in different fields, we will boldly outline our own, non-dogmatic definition, which is contextualized within research on everyday phenomena and routinized processes.[1] However, it also reflects the roots of the concepts.

First, Pike's innovation was to turn away from an epistemological debate over the objectivity and subjectivity of knowledge, instead turning towards a methodological solution. Similarly, we employ *etic* and *emic* as two methodological perspectives: *etic* refers to a researcher-analyst's outsider perspective, *emic* to a practitioner-informant's insider perspective. This definition has a weak spot: is etic knowledge not 'contaminated' by the researcher's own emic preconceptions (e.g. Perri and Bellamy, 2011: 233, 234; Yin, 2015: 16–18)? The issue was one of the key points of divergence between those primarily responsible for the concepts' spread, Kenneth L. Pike and Marvin Harris (see Harris, 1990: 49). This is a valid critique if one considers etic and emic as epistemic categories. Instead, we propose to apply them only with reference to data production methodology. Similarly, this limitation evades the (epistemological) argument that there is no guarantee that an informant will (or can) ever express their honest and unfiltered emic truth (for the discussion, see Olivier de Sardan, 2015: Chapter 3).

Second, a consequence of viewing these terms as a more or less value-laden dichotomy leads some researchers to suggest that it is possible to perform either an *emic* or an *etic* analysis, rather than understanding that they should be used in tandem. 'They were never intended to stand alone and so are inadequate when used in that way. Different information is gained from each form of analysis' (Hahn et al., 2011: 4). We do not view either etic or emic as superior (cf. Harris, 1980: Chapter 2) but rather see them as complementary methodological ways to approach the process or phenomenon under scrutiny.

Third, insiders normally are neither conscious of their emic system nor able to formulate it for the researcher: there is a difference between being *able to do* something within a culture competently and being *able to describe* it accurately (Hymes, 1970: 281–282). We therefore draw on the methodological literature in journalism studies (Berger, 2019; Flick,

2014; Reich, 2009) and related fields such as sociology (e.g. Dempsey, 2010; Larsen et al., 2008), applied linguistics (e.g. Haapanen, 2018; Perrin, 2013) and writing research (e.g. Grésillon and Perrin, 2014; Smagorinsky, 1994) and argue that an insider's view often benefits from being supported by etic data to gain an accurate emic point of view. The opposite also holds: the production of etic data can be informed and guided by emic knowledge.

Integration of the two perspectives

In order to evaluate the integration of the etic and emic perspectives in data production methodology, we have created a three-point taxonomy. A *low integration* methodology could be described as regular data triangulation. It means that the etic data sets, such as news items or observation notes, and the emic data sets, such as interview recordings, at least represent the same media field, but the data do not refer to the same instances of behaviour (e.g. the production of a particular news item). *Partial integration* refers to studies that produce etic and emic data sets that are closely related but that do not interact during the research process. The coverage of these data sets overlaps to some extent; for example, a researcher might analyse news items and interview the journalist who produced them. A *high integration* methodology involves using one type of data in order to elicit or analyse the other, as a result of which a data set is created that describes the instance of behaviour from both etic and emic perspectives; for example, a think-aloud protocol may be used to elicit (emic) work-process explanations from a journalist viewing an (etic) video record of their own work.

Gaining a more subtle understanding of the methodological solutions in our research corpus, we also categorize the studies in terms of etic-emic balance. Etic and emic data production methods may be in balance, or either one of them may be dominant. The balance has been assessed by using the following questions:

1. Does the methodology produce sufficient etic data to describe the behaviour relevant to the research question(s) that is explained by the emic data?
2. Does the methodology produce sufficient emic data to explain the behaviour relevant to the research question(s) that is described by the etic data?

If the answer to both questions is positive, the etic-emic data sets are in balance. If, on the other hand, the answer to question 1 is positive but to question 2 negative, there is an etic emphasis in the data collection. If the opposite is true, the emphasis is on the emic side.

To give just a few illustrations of the various combinations of integration and balance, we briefly present below five examples of studies from our corpus. To begin with a research framework with low integration and a balanced etic-emic approach, a paper authored by Tandoc and Skoric (2010) serves as a good example. They investigated the coverage of 'pseudo-events' in the Philippine press. Their etic data consisted of 2330 new articles from the leading Philippine newspaper. They were coded based on five categories. The emic data were produced by a questionnaire, which was answered by 100 Filipino journalists from 23 news organizations. The questionnaire was based on the same categories used in the content analysis, but the questions did not refer directly to those 2330 articles.

A partially integrated data production methodology with an etic emphasis is well illustrated by a paper authored by Ferrucci (2015). He studied how two news media, one weakly market-oriented and the other strongly market-oriented, covered the same news event, the 2013 St. Louis mayoral campaign. The etic data of this study consist of 60 news articles published in these media outlets. In order 'to generate more data to triangulate findings and gather further explanation' (Ferrucci, 2015: 200), the researcher also conducted interviews with three journalists. This formed the emic data of the study. We consider the integration of this study to be 'partial' because the interview procedure did not draw on the articles written by these particular journalists, but was 'only' informed by the three main themes that emerged in the text analysis. This study is categorized as having an etic emphasis because these three interviewees represent only a small proportion of the dozens of journalists who wrote the articles that were analysed.

The remaining three examples exemplify research frameworks with high etic-emic integration, although the balance between etic and emic varies in each case. For an etic emphasis, let us look at a paper authored by Harrison (2017), who studied the way the BBC deals with user-generated content (UGC) at its UGC hub. She spent 7 days observing 'whatever I wished' (Footnote 2) at the hub, and to complement this etic perspective with emic insights she was able to talk informally with all members of staff. For emic emphasis, we draw attention to the research framework designed by Grøndahl Larsen (2017). She studied how Norwegian journalists monitor, assess and make use of online information in investigative reporting of violent extremist groups. The emic data were constructed from 26 in-depth, face-to-face interviews with the journalists. The etic data serve as background for the interviews: the researcher acquainted herself with a broad selection of news articles, in-depth feature stories, journalistic television and book documentaries made by the interviewees and then asked the interviewees to describe and assess the production of these texts.

Finally, a research framework with high integration and a well-balanced etic-emic approach is exemplified by a paper authored by Van Hout and Jacobs (2008). Using an innovative method combining newsroom ethnography and computer-assisted writing process analysis, they explored how journalists transform press releases into news stories in the context of Belgian business news journalism. In their data production the etic and emic perspectives complemented each other perfectly: the data include fieldnotes from ethnographic observation, audio recordings of story meetings ($n = 53$) and interviews with reporters, copy-editors and the desk chief as well as internal memos, documents and emails. In addition, they utilized tools for recording keyboard strokes and mouse movements and for making screen capture videos of the observed writing processes. A video playback of the recorded writing process was then used as a stimulus in a retrospective interview ($n = 18$).

A systematic literature review

In order to investigate how the etic and emic perspectives have been combined in the study of journalistic practices, we carried out a systematic literature review.

Corpus production: Keyword search and manual filtering

First, we compiled a list of articles potentially relevant to the study. We searched articles from the complete World of Science database and two EBSCO databases, Academic Search Elite (ASE) and Communication and Mass Media Complete (CMMC). Our search included all texts that were categorized as having been peer reviewed and published in academic journals, as being written in English, published during the years 2000–2017, and conforming to the set of criteria defined below.

Each item had to include the search term journalis* combined with at least one of the following search terms: ethnograph*, observ* or interview*. After several trials with various search terms we chose these hypernym methodological keywords as they seemed to broadly cover the relevant methods and the words etic and emic themselves were rarely used. We concluded that this combination produced the most comprehensive corpus without including an unmanageable number of obviously irrelevant articles. The terms could be present in the article's title, abstract, or keywords. After removing duplicates through an automated duplicate recognition and manual confirmation, a corpus of 3018 articles remained.

In the second stage the corpus was split between the two authors (resulting in 1500 and 1518 references, respectively). We then manually filtered the references, mostly relying on their titles and abstracts, but also consulting their full texts when appropriate, so that they all met the following criteria:

1. *The language of the article is English.* English-language publications are the most influential, as they have a potential for the widest readership. In fact we also lack the competence to include many languages in our review.
2. *The genre is research article in a scholarly journal.* It could be argued that we should have included monographs, but it is likely that journal articles have also been published on research presented in monographs. Articles are therefore likely to provide a representative picture of the research done in the field.
3. *The research employs both emic and etic perspectives on the same material and/or mental, temporally distinct processes of journalistic production, and contains empirical data.* The focus of this article is on the practices of journalistic work, credible research on which requires empirical data.
4. *The focus is on professional journalism and not, for example, on journalism students' work.* Thus, the study describes established practices.

We processed our quotas independently, setting aside any articles that were difficult to classify. These cases we discussed and resolved together. This process of manual exclusion produced a remaining sample of 228 articles.[2] Intercoder reliability was assessed by cross-coding a random sample of 300 articles. The resulting Cohen's kappa (κ) value was 0.65, indicating acceptable reliability.

The final sample of 228 articles was again divided between the authors, closely read and coded manually by following our iteratively refined and completed coding scheme. Intercoder reliability was tested with a random sub-sample of 45 articles

(descriptive codes, like year of publication, were excluded from this analysis). For the most part, κ indicates acceptable reliability in recognition. However, the kappa for some values was 0 or 1 due to those values being exceedingly rare within the sub-sample. Difficulty in calculating intercoder reliability for rare conditions is an inherent weakness of Cohen's kappa (e.g. Vach, 2005), a weakness that may hide actual problems in the coding. For transparency, we will indicate the κ for each condition when reporting the results.

Corpus analysis: Categorization

We coded the remaining sample ($n=228$) using a constant comparison protocol: when the need to add new codes or categories arose, we consulted each other and jointly decided on the addition, its definition, and the possible need for retroactive corrections to previously coded items.

The first stage was to collect descriptive data for each article, including its name, author(s), keywords, the name of the publishing journal, year of publication and the study's target country. Then, further methodological information was collected, of which we detail below the information that we address in the results of this literature review.

Data source(s). By this we mean the various resources used in producing the data for analysis, for example *published news items* (for content analysis) or *journalists* (for interviews). We identified and coded for 29 types of resources, including various types of journalistic material, document sources and human actors. Coding this category proved problematic in light of the intercoder reliability check, with only 22 values producing $\kappa > 0.6$. However, all the most common resource types (with five or more appearances in the sub-sample used in the intercoder reliability check) were coded with an acceptable degree of accuracy ($\kappa \geqslant 0.69$).

Platform. We coded each article by the medium or media it sought to examine. Most studies were easily categorized as involving *newspapers* ($\kappa=0.86$), *magazines* ($\kappa=0.65$), *television* ($\kappa=1$), *radio* ($\kappa=0.69$) or *online only* ($\kappa=0.78$). We also acknowledged the (uncommon) platform of *books* ($\kappa=1$). The accuracy of the coding was poor with regard to the (uncommon) codes *unclear* ($\kappa=0.48$) and *multi-platform* ($\kappa=0.37$). It is worth mentioning that the weak kappa values here and elsewhere are probably related to the fact that the description of both the data and the methods was vague and incomplete in many articles, a matter to which we will return in the Discussion.

Focus. We identified 21 phases or parts of the journalistic process that the studies focused on. This included salient processes such as *interviewing* or *sourcing*, but also more abstract points of departure such as *ethical considerations* or *management*. This category, too, suffered from poor intercoder reliability, with only 15 conditions being recognized with $\kappa > 0.6$ accuracy. The inaccuracies are confined to uncommon research foci (e.g. *social media use*), and all the most common codes (appearing more than five times in the sub-sample) were recognized with an accuracy of $\kappa \geqslant 0.64$.

Etic-emic emphasis. We coded each article based on how much it emphasized either etic ($\kappa = 0.64$) or emic ($\kappa = 0.67$) data. If the two methods appeared balanced, we coded it as such ($\kappa = 0.66$).

Etic-emic integration. Each article was also coded according to the degree to which it integrated etic and emic data production. The integration was defined to be either *high* ($\kappa = 0.56$), *partial* ($\kappa = 0.64$) or *low* ($\kappa = 0.63$), as explained in the previous section.

Media type. We also attempted to code each article by the kind of news organisation(s) it investigated (e.g. *professional legacy media* or *professional new media*). However, the results of intercoder reliability test for this category were poor, and we will report on the media types only briefly.

Results: Emphases and research gaps

In this section, we present the findings of our literature review. We will start with a description of the studies in our corpus and move on to their content, focus and the integration of, and balance between, etic and emic perspectives in their data production. It is worth emphasizing that we did not evaluate the quality of the studies in any way, for example in terms of the credibility or importance of the results.

Description of the data

Our sample of 228 articles was unevenly distributed by publication year: two of the articles were published in 2000, while 37 came out in 2017. The increase was relatively steady, but it is unclear just from this sample whether it is proportional to a general increase in scholarly publishing or if it indicates an increased appreciation of the etic-emic combination methodology.

We limited our corpus to journal articles written in English, with the unsurprising result that the geographical distribution of our data is biased towards the United States (40 articles, 17.5% of the total) and the United Kingdom (31, 13.6%); after these two countries came Belgium (14, 6.1%), Spain (14, 6.1%), The Netherlands (10, 4.4%), Slovenia (10, 4.4%), Finland (9, 4.0%) and Germany (9, 4.0%) in terms of number of articles. It should be noted, however, that in the case of many countries, several articles deal with data produced in the same project. Overall, the geographical distribution of articles spans 44 countries from all the inhabited continents, although the distribution is not very even: 140 (61.4%) articles deal with the mediascape in European countries, 45 (19.7%) in North America, 23 (10.0%) in Asia (incl. Russia 1), 14 (6.1%) in South and Central America, 12 (5.3%) in Africa and 5 (2.2%) in Australia. In addition, two articles have a global coverage, two deal with English-speaking countries in general and a few deal with more than one country. There were also some articles in which the geographical context remains unclear and which were therefore not coded for any country.

As for publication venues, three journals dominate: *Journalism Practice* (16.6%), *Journalism* (13.9%) and *Journalism Studies* (12.8%). In total, our sample includes articles

from 68 journals, but the drop after the top three journals is significant: the publication with the fourth-highest number is *Digital Journalism*, with a 3.7% share.

The articles credit 329 scholars as authors, and only a few of these stand out in terms of number. Only 35 authors are credited with more than one publication. Data from one study were often used in several publications, with the result that no researcher can be considered to have been a particularly avid user of this data collection methodology.

Exactly 797 different author-supplied keywords were used in the sample, including what are, to all intents and purposes, duplicates (e.g. *newspaper* and *newspapers*) and proper names (e.g. *Al Jazeera*). Discounting obvious (and very common) keywords such as *journalism* and *news*, the sample's keywords are very diverse: only 26 appear more than ten times. Of the top keywords only a few are prolific enough to warrant a mention: *newspaper* and *newspapers* (combined 5.3%), *news value* and *news values* (combined 4%) and *sources* (2.9%).

Objects of interest

Research appears to have overwhelmingly focused on *professional legacy media*, such as traditional print newspapers or broadcast television: legacy media was the object of study in over 86% of the articles we reviewed. However, the intercoder reliability for recognizing legacy media was weak ($\kappa = 0.29$). We performed a reliability check with a sub-sample of 45 articles, and were in agreement that 40 of them involved professional legacy media and one did *not* involve legacy media. We disagreed on the remaining four articles, which in this small sample was enough to bring down the κ. This disagreement notwithstanding, we can confidently report that the vast majority of the research in our sample focuses on professional legacy media.

The studies are also quite conservative with regard to the type of publishing platform. The vast majority (56.6%) of articles reported studies on *newspapers* ($\kappa = 0.86$), which was also reflected in the keyword statistics. Following it were studies on *television* (25.4%, $\kappa = 1$) and *online only* (20.2%, $\kappa = 0.78$) news organisations. *Radio* was covered in 9.2% of the articles ($\kappa = 0.69$), while only 5.2% of them investigated *magazines* ($\kappa = 0.65$). Some 3.9% of the articles were coded as focusing on *multi-platform* organizations. However, the intercoder reliability for this latter code was weak ($\kappa = 0.37$), and we suggest dismissing the result as unreliable.

Methods and material

We identified 28 different types of data sources. As far as human sources are concerned, the most common were, by far, rank-and-file *journalists* (84.7%, $\kappa = 0.63$), followed by *editors and sub-editors* (55.7%, $\kappa = 0.69$), *entire news organisations* (25.4%, $\kappa = 0.9$), and *non-journalists* such as PR specialists, politicians and expert-interviewees (22.8%, $\kappa = 0.81$). As for material sources, the top four were *written news items* (43.9%, $\kappa = 0.82$), *audio-visual news items* (15.8%, $\kappa = 1$), news organisations' *internal documents* such as reports and statistics on readership and finances (7.9%, $\kappa = 0.65$) and various *social media posts* (7.9%, $\kappa = 0.66$). The superiority of written sources over audio-visual (let alone *audio news items*, which accounted for only 3.5%) may be due to the strong

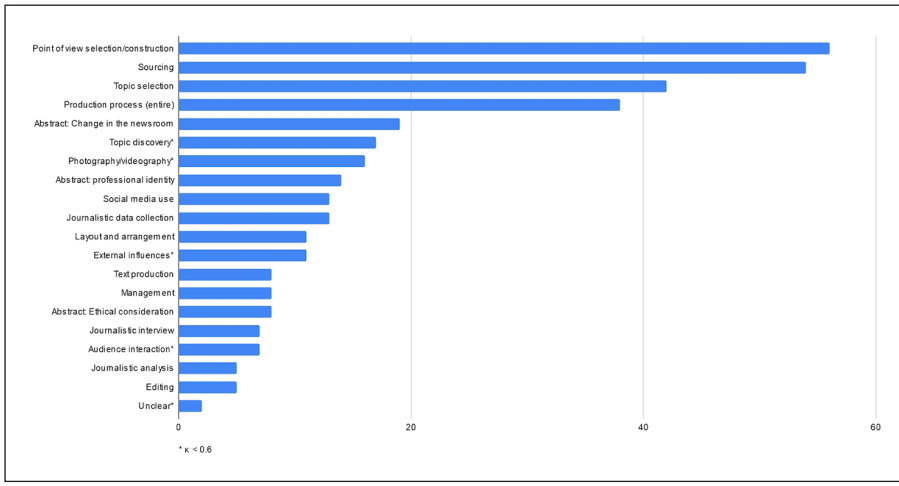


Figure 1. The part of the journalistic work process in focus.

tradition of newspaper research and the fact that a data corpus of written texts is easier to compile, handle and analyse than audio-visual content. This is a mismatch that needs to be addressed when planning new research.

It is also worth pointing out the seemingly neglected areas of investigation: *photographs and videos* were studied in only 4.8% of articles (κ = 1). *Photojournalists* themselves, too, appear to have been overlooked (6.1%), but due to the low intercoder reliability of this code (κ = 0.48) the finding is only tentative.

Research focus

We identified 20 distinct aspects of the journalistic process on which the articles in our sample focused (see Figure 1). Three stood out as most common: *Point of view selection* (24.6%, κ = 0.64), *sourcing* (23.7%, κ = 0.8) and *topic selection* (18.4%, κ = 0.72). A considerable number of studies also aimed to look at the *production process in its entirety*, with no specific focus (16.7%, κ = 0.69). We will weigh the implications of these choices further below, in the Discussion. Apart from these four codes, scholarly interest seems to be evenly distributed among the remaining foci. While all the other aspects of the journalistic process were almost equally ignored, some of these omissions are egregious enough to be discussed a little further. The topics that we mention next are, in our opinion, just as salient and important to the journalistic end product as the four topics that we identified as being most studied.

Photography and videography attracted only limited attention (7%, κ = 0). The *editing process*, too, aroused scant interest: 2.2% of the articles in our sample (i.e. 5 studies) focused on how sub-editors give journalistic items their final polish (κ = 1). Similarly, *management* (3.5%, κ = 1), *layout and arrangement* (4.8%, κ = 0.85) and *social media use* (5.7%, κ = 0) seem to have been of little interest. Although some of the kappa values

Table 1. Integration and balance of the data collection.

	Emic-emphasis ($n = 40$)	Well-balanced ($n = 155$)	Etic-emphasis ($n = 33$)
high integration ($n = 50$)	4.4% ($n = 10$)	16.2% ($n = 37$)	1.3% ($n = 3$)
partial integration ($n = 123$)	7.0% ($n = 16$)	39.5% ($n = 90$)	7.5% ($n = 17$)
low integration ($n = 55$)	6.1% ($n = 14$)	12.3% ($n = 28$)	5.7% ($n = 13$)

were undermined by a few disagreements in the 45-article sub-sample used to calculate intercoder reliability, 44 articles were unanimously identified as *not* involving social media use and 43 as *not* focusing on photography or videography. Thus, we consider these results to be indicative, if not accurate.

Integration and balance of the data

As explained above, we used a 3-point scale to assess both the integration and the balance of the etic and emic perspectives. In this way, we were able to define nine different combinations of perspectives, as shown in Table 1.

In line with the reasoning presented in the introductory section, it might be tempting to interpret studies estimated as 'high integration' and 'balanced' as of the highest quality; 37 out of the total number of 228 studies (16.2%) fell into this category. However, as stated above, this is not our intention.

Looking at the whole, we see that the only clearly larger category is the middle square of the 9-square field: In 90 of the 228 studies (39.5%), the data collection methodology was assessed as being balanced and partially integrated. This distribution is partly explained by the fact that the middle category became something of a 'catch-all category': many studies described their methodology so vaguely that we had difficulty in confidently placing them at either end of the continuum. While this undermines the accuracy of our review, it is an important result in its own right: data production methods are often insufficiently described, which makes it difficult for readers – and referees – to properly assess the study's quality.

It may also be the case that few studies are expressly designed with an etic-emic-combination methodology in mind. Rather, the combination may emerge by chance through *ad hoc* adjustments or additions born from shortcomings or opportunities discovered during the research.

Discussion: Methodological descriptions must be improved

We reviewed thousands of research articles reporting on studies into journalism. Our goal was to explore the use of a particular methodological combination: that of emic and etic data production. This combination, in our view, is extremely helpful in describing and explaining processes that may be esoteric or heuristic and therefore difficult to grasp accurately through simple means. Many parts of journalistic work, we argue, fit this description. However, only some 7.6% of our complete corpus, that is 228 articles, did

employ both etic and emic methods of data collection – and fewer still did so with the kind of rigour we would advise.

There is one compelling reason why the etic-emic combination is so far less popular than the more straightforward data collection methodologies: it is laborious. Using different kinds of data sets and different production methods will introduce analytical complexity – and make it difficult to report the study concisely and clearly. Integrating etic and emic data production methods will also introduce practical complexity, as these procedures must be scheduled with a particular interwoven sequence in mind. This sequencing will often require more time and thus increase costs. Another practical obstacle might be the fact that newsrooms are not always easily accessible to researchers, just as journalists are not always willing to open up their work.

Another reason for this small number may be academics' inclination to divide research into as many separate publications as possible. This so-called *salami slicing* can lose the holistic view, reducing the explanatory value of the project. It should be noted that a highly integrated etic-emic-research study is not divisible into separate presentations. Thus, the small number of high-integration articles ($n = 50$) does reflect the genuine scarcity of the methodology; it obviously cannot be explained by scholars spreading out etic and emic parts of their research over different publications.

Of the 228 articles that benefited from etic and emic data production, many investigated the first steps in the editorial process: *topic discovery* (17 articles), *topic selection* (42 articles), *point of view selection* (56 articles) and *sourcing* (54 articles). There is overlap in these categories (i.e. several of these foci occur in the coding of the same article), but they clearly form a distinct group of studies that is an order of magnitude larger than studies with other types of foci. Arguably, these processes are the ones most difficult to reliably circumscribe by simpler means: omitted sources, rejected viewpoints and editorial debates could hardly be captured by emic or etic data alone. At the same time, these factors are closely related to some of journalism studies' key interests: Who gets their voice heard? How is the news framed? What news ends up on the public agenda? These are questions of representation, fairness and objectivity.

The emphasis on the beginning of the editorial process can be considered disproportionate. It emphasizes the agency of the individual reporter and their intentions. It undervalues the impact, say, a sub-editor or a newsroom manager has on the output. Therefore in order to accurately describe and holistically understand the journalistic work process, we recommend that researchers branch out in their application of the etic-emic combination methodology. Similarly, the scholarship is skewed in favour of *professional legacy media*, while leaning away from *professional new media*, *news aggregators*, *freelancers* and *professional amateurs*. In terms of resources used in the analysis, studies are disproportionately reliant on *written sources* as opposed to *audio-visual material*. All this reflects a kind of conservatism, despite the fact that our literature review has looked specifically at a data production methodology that can be considered innovative.

We also looked at how the authors of these 228 articles evaluated their data production methodologies. To put it briefly, shortcomings and challenges were not raised very often (Boesman et al., 2015; Perrin, 2011; Reich, 2010; Vandevoordt, 2016, 2017); the main focus was on how combining the perspectives can add value. Even there, discussion was rather thin and general. Taken together, however, these (mostly) fragmented

mentions build a good picture of what can be achieved by integrating the etic and emic perspectives.

The added value of triangulation is principally discussed in relation to the interview method. Firstly, the authors suggest that supplementing interviews with etic stimuli yields a broader picture and a deeper understanding (e.g. Richardson and Barkho, 2009; Van Hout et al., 2011). Retrospective verbalization may benefit from the use of stimuli by also bringing into consideration details that might otherwise be neglected due to memory limitations (e.g. Larsson et al., 2017; Reich, 2008, 2010) and the fact that routine work tasks may have left only vague traces of memories (Manninen, 2017). Furthermore, grounding the interview in real-life details keeps the discussion from slipping onto a general, abstract or hypothetical level (e.g. Boesman et al., 2015; Brüggemann, 2012; Vandevoordt, 2016, 2017) and gives the researcher what seems to be first-hand experience of the work process (Mohd Shariff, 2011; Niblock and Machin, 2007).

Secondly, the researcher can use pre-produced etic data, on the one hand, to verify that the interviewee's account is accurate (e.g. Coddington, 2012; Kuang, 2017), or that the account matches actual work practices (e.g. Alkaff and McLellan, 2017; Coddington, 2012; Hellman and Jaakkola, 2011; Hermans et al., 2014; Kuang, 2017; Lavie and Lehman-Wilzig, 2003). On the other hand, supplementing the interviews by presenting etic findings (Poler-Kovačič and Erjavec, 2010; Vandevoordt, 2016), or at least by making the interviewees aware that their work has been observed (Godler and Reich, 2013), is likely to encourage the interviewee to speak the truth without embellishing it. Conversely, data collected from an etic perspective can be verified through interviews (e.g. Erjavec, 2004; Poler-Kovačič and Erjavec, 2010; Vertommen et al., 2012).

All in all, the added value of combining the two perspectives is discussed only in a little over 40 of the 228 articles, and then most often only briefly, in a sentence or two. We also made an even more unfortunate discovery, which applies not only to these 228 articles but also to the 3018 articles that we analysed in the second stage of creating our corpus: descriptions of the research methods and the data sets they have produced are often unacceptably lax, superficial and incomplete. Thus, it has to be said that authors, editorial staff and referees should pay far more attention to methodology, even if the study is labelled as 'ethnographic'.

We do not intend this review to be taken as advocating English over other languages, journal articles over books, balanced data distribution over etic or emic emphases. On the contrary, we are in favour of producing research in different languages and platforms, with different approaches and areas of interest. However, we do want to make a statement in favour of well-planned research, careful methodological description and widely inclusive research interests.

Conclusion: Effort will be rewarded

In this paper, we have discussed the methodology of data production in research on journalistic work processes and have suggested that a combination of an outsider's etic and an insider's emic perspectives can produce higher quality and more comprehensive data. We have presented the results of a broad research literature review and discussed the rarity of this methodology, which may be partly justified by how labour intensive it is.

We conclude by asking a fundamental question: why do research? Reasons vary, of course, but when it comes to a field of research with such a close connection to a particular profession as in journalism studies, a key motive is undoubtedly the wish to improve professional practices and outcomes.

In a famous phrase, psychologist Lewin (1948) suggested that 'if you want truly to understand something, try to change it'. In the pursuit of developing work processes, we want to turn the phrase around: *If you want to change something, try to understand it*. This is exactly what combining perspectives makes possible: asking and answering both descriptive *what*-questions and explanatory *why*-questions.

As our literature review has shown, conducting such research is not very common in journalism studies, and there is room for improvement especially in integrating the etic and emic perspectives. Such research set-ups are challenging to carry out and demanding in terms of both time and human resources. However, we believe that the investment is well worth while in the long run. As Daniel Perrin once told us, 'methodological complexity is a good predictor of research gaps'.[3]

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD


Lauri Haapanen https://orcid.org/0000-0002-1973-4843

Supplemental material

Supplemental material for this article is available online.

Notes

1. The methodology we present can be seen to reflect the same procedures and goals as so-called Mixed Method Research which, however, adheres programmatically to the idea that both quantitative and qualitative methods should be involved (e.g. Creswell, 2011).
2. A list of the 228 articles in this corpus can be found as Supplemental material.
3. The lecture given by Professor of Applied Linguistics D. Perrin was held in the University of Jyväskylä, Finland, on January 18, 2019.

References

Alkaff S and McLellan J (2017) Same news, different stances? A comparative media discourse investigation of hard news texts in the New Straits Times and Berita Harian. *Pertanika Joxurnal of Social Science and Humanities* 25(2): 511–539.

Berger AA (2019) *Media and Communication Research Methods. An introduction to qualitative and quantitative approaches*. London: Sage.

Boesman J, d'Haenens L and Van Gorp B (2015) Triggering the news story. *Journalism Studies* 16(6): 904–922.

Brüggemann M (2012) Transnational trigger constellations: Reconstructing the story behind the story. *Journalism* 14(3): 401–418.

Coddington M (2012) Building frames link by link: The linking practices of blogs and news sites. *International Journal of Communication* 6(20): 2007–2026.

Crane JG and Angrosino MV (1984) *Field Projects in Anthropology: A Student Handbook*. Prospect Heights, IL: Waveland Press.

Creswell JW (2011) Controversies in mixed methods research. In: Denzin NK and Lincoln YS (eds) *The SAGE Handbook of Qualitative Research*. London: SAGE, pp.269–284.

Erjavec K (2004) Beyond advertising and journalism: Hybrid promotional news discourse. *Discourse & Society* 15(5): 553–578.

Dempsey NP (2010) Stimulated recall interviews in ethnography. *Qualitative Sociology* 33: 349–367.

Ferrucci P (2015) Primary differences: How market orientation can influence content. *Journal of Media Practice* 16(3): 195–210.

Flick U (2014) *An Introduction to Qualitative Research*. London: Sage.

Godler Y and Reich Z (2013) How journalists "realize" facts. *Journalism Practice* 7(6): 674–689.

Grésillon A and Perrin P (2014) Methodology: From speaking about writing to tracking text production. In: Jacobs E-M and Perrin D (eds) *Handbook of Writing and Text Production*. Berlin: De Gruyter Mouton, pp.79–111.

Grøndahl Larsen A (2017) Investigative reporting in the networked media environment. *Journalism Practice* 11(10): 1231–1245.

Haapanen L (2017). *Quoting practices in written journalism*. PhD Thesis, University of Helsinki. Available at: helda.helsinki.fi/handle/10138/174618 (accessed 9 May 2021).

Haapanen L (2018) Overcoming methodological challenges in text production research: A holistic approach through data triangulation. *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 11: 45–70.

Hahn C, Jorgenson J and Leeds-Hurwitz W (2011) A curious mixture of passion and reserve: Understanding the etic/emic distinction. *Éducation et didactique* 5(3): 145–154.

Harris M (1990) Emics and etics revisited. In: Headland TN, Pike KL and Harris M (eds) *Emics and Etics: The Insider/Outsider Debate*. London: Sage, pp.48–61.

Harris M (1980) *Cultural Materialism. The Struggle for a science of culture*. New York: Vintage Books.

Harrison JA (2017) Rust belt boomerang: The pull of place in moving back to a legacy city. *City & Community* 16(3): 263–283.

Headland TN, Pike KL and Harris M (eds) (1990) *Emics and Etics: The Insider/Outsider Debate*. London: Sage.

Hellman H and Jaakkola M (2011) From aesthetes to reporters: The paradigm shift in arts journalism in Finland. *Journalism* 13(6): 783–801.

Hermans L, Schaap G and Bardoel J (2014) Re-establishing the relationship with the public. *Journalism Studies* 15(5): 642–654.

Hymes DH (1970) Linguistic method in ethnography: Its development in the United States. In: Garvin P (ed) *Method and Theory in Linguistics*. The Hague: Mouton.

Janićijević N (2015) The hybrid approach in the research of organizational culture. In: *Proceedings of the 14th European Conference on Research Methodology for Business and Management Studies*. Valletta, Malta, 11−12 June, pp. 314−321. Reading, UK: Academic Conferences and Publishing International Limited.

Koliska M (2015) *Transparency and trust in journalism: An examination of values, practices and effects*. PhD Thesis, University of Maryland, College Park.

Kuang X (2017) Negotiating the boundaries of news reporting: Journalists' strategies to gain access to and report on political information in china. *MedieKultur: Journal of Media & Communication Research* 33(62): 35–51.

Larsen D, Flesaker K and Stege R (2008) Qualitative interviewing using interpersonal process recall: Investigating internal experiences during professional-client conversations. *International Journal of Qualitative Methods* 7(1): 18–37.

Larsson AO, Kalsnes B and Christensen C (2017) Elite in interaction. Public service broadcasters' use of twitter during national elections in Norway and Sweden. *Journalism Practice* 11(9): 1137–1157.

Lavie A and Lehman-Wilzig S (2003) Whose news? goes gender determine the editorial product? *European Journal of Communication* 18(1): 5–29.

Lewin K (1948) *Resolving Social Conflicts. Selected Papers on Group Dynamics*. New York: Harper & Brothers.

Manninen VJE (2017) Sourcing practices in online journalism: An ethnographic study of the formation of trust in and the use of journalistic sources. *Journal of Media Practice* 18(2): 212–228.

Mohd Shariff SZ (2011) Local broadcast journalists: Journalistic autonomy and level of job perception in their workplace. *Human Communication* 14(3): 259–283.

Niblock S and Machin D (2007) News values for consumer groups: The case of independent radio news. *Journalism* 8(2): 184–204.

Olivier de Sardan J-P (2015) *Epistemology, Fieldwork, and Anthropology*. New York: Palgrave Macmillan.

Perri 6 and Bellamy C (2011) *Principles of Methodology: Research Design in Social Science*. London: Sage.

Perrin D (2011) "There are two different stories to tell". Collaborative text-picture production strategies of TV journalists. *Journal of Pragmatics* 43(7): 1865–1875.

Perrin D (2013) *The Linguistics of Newswriting*. Amsterdam: John Benjamins.

Pike KL (1954) *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Dallas, TX: Summer Institute of Linguistics.

Pike KL (1988) Cultural relativism in relation to constraints on world view–An emic perspective. *Bulletin of the Institute of History and Philology* 59(2): 385–399.

Pike KL (1990) On the emics and etics of Pike and Harris. In: Headland T, Pike KL and Marvin H (eds) *Emics and Etics: The Insider/Outsider Debate*. London: Sage.

Poler-Kovačič M and Erjavec K (2010) Production process of unlabeled advertorials in the Slovenian press. *Communications-European Journal of Communication Research* 35(4): 375–395.

Reich Z (2008) The roles of communication technology in obtaining news: Staying close to distant sources. *Journalism & Mass Communication Quarterly* 85(3): 625–646.

Reich Z (2009) *Sourcing the News. Key Issues in Journalism–An Innovative Study of the Israeli Press*. New York: Hampton Press.

Reich Z (2010) Measuring the impact of PR on published news in increasingly fragmented news environments. *Journalism Studies* 11(6): 799–816.

Richardson JE and Barkho L (2009) Reporting Israel/Palestine. Ethnographic insights into the verbal and visual rhetoric of BBC journalism. *Journalism Studies* 10(5): 594–622.

Smagorinsky P (ed) (1994) *Speaking about Writing: Reflections on Research Methodology*. Thousand Oaks: Sage.

Tandoc EC Jr and Skoric MM (2010) The pseudo-events paradox: How pseudo-events flood the Philippine press and why journalists don't recognize it. *Asian Journal of Communication* 20(1): 33–50.

Vach W (2005) The dependence of Cohen's kappa on the prevalence does not matter. *Journal of clinical epidemiology* 58(7): 655–661.

Van Hout T and Jacobs G (2008) News production theory and practice: Fieldwork notes on power, interaction and agency. *Pragmatics* 18(1): 59–85.

Van Hout T, Pander Maat H and De Preter W (2011) Writing from news sources: The case of apple TV. *Journal of Pragmatics* 43(7): 1876–1889.

Vandevoordt R (2016) Covering the Syrian conflict: How Middle East reporters deal with challenging situations. *Media, War & Conflict* 9(3): 306–324.
Vandevoordt R (2017) Why journalists covered Syria the way they did: On the role of economic, social and cultural capital. *Journalism* 18(5): 609–625.
Vertommen B, Vandendaele A and Van Praet E (2012) Towards a multidimensional approach to journalistic stance. Analyzing foreign media coverage of Belgium. *Discourse Context & Media* 1(2–3): 123–134.
Yin RK (2015) *Qualitative Research from Start to Finish*, 2nd edn. New York: Guilford.

Author biographies

Lauri Haapanen, PhD is a University Lecturer at the University of Jyväskylä, Finland. His current research investigates journalists' work practices from various angles, for example, how they participate in, take into account, and are influenced by social media when writing their media items. Haapanen received his Ph.D. from the University of Helsinki (2017). His article-based dissertation examined quoting practices in written journalism, and its key results are also widely discussed in two major handbooks: *The Routledge Handbook of Language and Media* (2017, a chapter *Media and Quoting*, co-authored with Daniel Perrin) and *The International Encyclopedia of Journalism Studies* (Wiley 2019, an entry *Reporting and Interviewing*, co-authored with Mats Nylund).

Ville JE Manninen, PhD is a Researcher at the University of Vaasa's InnoLab research platform. He has also worked as a researcher and a lecturer at the University of Jyväskylä, focusing on journalistic work practices and media plurality. Since 2015 he has been involved with the EU-funded and EU-wide Media Pluralism Monitor (MPM) project, and in 2018 joined a nation-wide consortium of Finnish media scholars to report on media plurality for the Ministry of Transport and Communications. Manninen is a journalism studies graduate and has working experience from seven newspapers.