

Coping with Epistemological Restrictions in Interdisciplinary Emotion Research

Elisabeth Huber, *Institute of Social and Cultural Anthropology*
Freie Universität Berlin, Germany
e.huber@fu-berlin.de

Abstract

Based on the findings of an EU Horizon 2020 funded project investigating Knowledge Complexity and bias in (big) data (K-PLEX: www.kplex-project.eu), this article focuses on the validity of 'emotion data'. By means of a series of in-depth interviews and an online survey, researchers in various disciplines were asked what data they are in a position to validate and how they adjust to methodologically induced knowledge gaps. Social scientists use a multitude of qualitative and quantitative methods to explore emotions and affects of individual persons, groups, and even entire societies. Epistemological questions thus arise as to whether valid data can be obtained at any scale. The longstanding issue of sample representativeness in sociology re-emerges in data-driven approaches yielding big data on emotions and affects. Closely connected to the matter of scope, this article explores the generalizability of research findings. Framed as the WEIRD-problem of psychological data created in laboratory settings, the criterion of external validity can also be applied to the question of cross-validation with a mixed method design. Especially with regard to the sharing and integration of 'emotion data', the consequences of information loss through datafication are addressed. From a critical epistemological perspective, validity is finally discussed in terms of reflexivity fostered by participative research designs. As communicative validation of data and interpretations with research subjects can entail both methodological and ethical challenges, researchers' responses to the limitations in establishing credibility are also analysed.

Introduction

In the humanities and social sciences, discussions on what constitutes good scientific practice date back to the beginnings of empirical research. With the differentiation and specialization of the disciplines, some of the principles have shifted in focus, but some controversies still persist. The tension between science and art falls into this ongoing discussion. Do scholars in the social sciences need to follow strict criteria of demonstrating and justifying their decisions within the research process, or should they be granted the freedom to artfully create an oeuvre that also meets aesthetic demands? Another fundamental debate centres on the conflict between the positivist criteria of accuracy and the criteria of accountability prevalent in critical ontological debates.

This article investigates how scholars from the social sciences, undertaking research on emotions and affects, encounter questions related to validity. What criteria do they regard as important? How should other researchers be convinced to trust the data and the conclusions drawn from them? And what kind or degree of validity results from data constructed by the researchers (they themselves being an 'instrument' in observing and listening) in contrast to data yielded by machines and algorithms? When researchers identify the limits and shortcomings of validity claims, how do they circumnavigate these limits?

Research on emotions and affects is particularly well-suited to an investigation of the various dimensions of validity, since the object of research manifests itself both explicitly and implicitly in a variety of phenomena. For instance, one can point to the role of the emotions and affects in decision-making processes or their shaping of cultural-historical practices and differences. One can even use applied research in the computational modelling of emotions. Based on primary data collected via 15 in-depth interviews and a survey with 123 participants from quantitative and qualitative backgrounds, this article seeks to highlight some of the current challenges and strategies in conducting solid research and reporting on it convincingly. Part of an EU Horizon 2020 funded project entitled Knowledge Complexity (K-PLEX), the interdisciplinary team located at the Freie Universität Berlin focused on epistemic cultures and issues of datafication and information loss among scholars carrying out research on emotions and affects.

Dimensions of Validity in Emotion Research

Validity in social-scientific research has been discussed in a range of ways. On the one hand, the concept of validity is associated with choosing the right methods in order to collect accurate data. On the other, it concerns data analysis, that is, the process of evaluating the data collected and their potential for meaning-making. The third dimension of validity, which complements both descriptive and interpretive validity, is

theoretical validity. The latter pertains to the applicability of theoretical statements to a certain range of contexts. In quantitative approaches, most attention is drawn to the accuracy of measurement. Researchers are preoccupied with the question of whether their approaches actually measure what they want to measure. A multitude of labels such as construct validity, content validity or internal validity have been offered in response to methodological claims to validity. Attempts to translate the validity criteria for qualitative approaches have been undertaken by several researchers, e.g. Margaret LeCompte and Judith Goetz (1982). In qualitative research, the ideas behind these labels are usually well-known but seldom articulated explicitly. In contrast to quantitative approaches, researchers face greater difficulties in showing how the evidence contributes or relates to the theory put forward.

More recent propositions differentiate between first-order criteria such as credibility, authenticity, criticality, and integrity, and second-order criteria such as explicitness, vividness, creativity, thoroughness, congruence, and sensitivity (Whittemore, Chase and Mandle, 2001). The ranking suggests that once researchers have made sure that they comply with the primordial principles of transparency and confirmability, they can pay attention to more fluid and flexible criteria. Especially for researchers who embrace a perspective that foregrounds the emic aspects, this ranking is problematic. Scholars in the field of emotion research who participated in the K-PLEX project formulated the challenge of representing the internal perspectives of research subjects in two respects: First, capturing collective emotional experiences in historical subjects was associated with high degrees of uncertainty; Second, defining distinct emotions in linguistic terms and in that sense referring to learned cultural concepts, rather than felt or sensed inner states, was identified as potentially distorting and simplifying.

If 'emotion, like poetry, gets lost in translation' (Beatty, 2019: 10), providing supplementary contextual information becomes indispensable. Social scientists unanimously agreed that the transferability of the meaning of words uttered in a certain context was highly problematic. As a way of comparing and translating, this is related to the question of external validation. Depending on the research object and the methodological approach, different answers have been given to the question of what additional contextual information is necessary to judge the validity of data and conclusions in emotion research (cf. Lehmann and Huber, 2018).

Some researchers were chiefly preoccupied with the validity of conclusions. Paul Ekman's attempt to compare emotions in a cross-cultural way and to identify a set of universally valid basic emotions was mentioned as an example. Validity was found to be only temporary, as Ekman himself had to revise the list of basic emotions several times. Although the data were not collected in laboratory settings, the dialogic interactions with research participants with the aim to induce emotions were criticized as artificially constructed situations that did not correspond with locally established speech-event forms. Generalization for real-life situations could therefore not be assumed. Moreover, some linguists and psychologists pointed to the fact that people do not behave in ways that can be so neatly categorized, as complex interactional behaviour patterns and ritualized interaction strategies demonstrate. The analysis of emotions must not be reduced to controlled experimental situations that exclude any alternative explanations, or one accepts results that are no longer valid. Cross-checking in interviews was then suggested as a strategy for verification. By means of research subjects' self-assessment, validity turns into validation and intersubjective agreements can be reached. Potential difficulties encountered – be it comprehensibility, memory or social desirability, or more fundamental ethical questions regarding the possibility that researchers and research participants might not necessarily pursue the same interests, commitments, and goals (Sandelowski, 1993) – have been touched on but not explored in detail.

Even if many researchers agree that only detailed narrative accounts can do full justice to the complexity of emotions, how to establish validity in narrative research remains a controversial issue (cf. Creswell, 2013). In the interviews and survey of the K-PLEX project, anthropologists often pointed to the provenance and context of origin and assured validity by taking protocols and reflecting on how the data were gathered. A sociologist working on the relationship between scandalization, moralization, and emotions by conducting qualitative analyses of social media content went so far as to say that the meaning of emotions depended entirely on the context. Computational social scientists, however, deplored the fact that emotion recognition in speech analysis based on machine learning was exclusively relying on predefined categories and examples. As the machine cannot deal with exceptions or context, researchers need to cope with the majority bias in identifying emotions. This point is well illustrated by the statement: 'We know that it is only valid if there is

one really prototypical situation'. A coping strategy intended to compensate for this problem was to be found in the collection of more 'noisy' data. However, this may give rise to other challenges. In a data-poor environment that consists mostly of noise, missing data can lead to falsely rejecting a valid hypothesis. Using data yielded by instruments developed at the initial stage of a research project and data from refined instruments in later stages clearly affects comparability as these data must be evaluated differently.

With big data, the question of validity is raised in new ways, in terms of both the data quality and the conclusions that can be drawn from the analyses. In big data research, less deterministic approaches referring to patterns instead of essentialist categories and classifications might alleviate some of the traditional social scientists' challenges. In addition, non-intrusive methods can yield more valid results as social desirability and interviewer effects become insignificant. Yet, other criteria such as sample representativeness are not met, since some groups do not leave data traces. Are these findings nevertheless valid? Referring to the second-order criteria mentioned above, I want to reiterate that methodological and theoretical rigour is not the only means through which validity can be achieved. Patterns detected do not necessarily reflect what is in the external world, but they often build upon the researchers' conceptualizations and creative imaginations. This is not to say that researchers invent data unrelated to phenomena and events, but reducing scientific practice to blindly clinging to data is not considered professional mastery. Big data research reformulates the relationship between validity and creativity in new ways. Data-driven approaches imply a high degree of serendipity as analyses produce findings that were not actively searched for. Big data analyses conducted by the interview partners in the K-PLEX project had found correlations between the Tsunami in the Indian Ocean in 2004 and the first recognizable wave of hate speech in Finnish social media. A second example pertains to German political parties on Facebook and the different emojis associated with each party. The right-wing party had received most of the angry faces while the social democrats scored higher with hearts. Interpreting the statistical correlations between political parties and emojis is not possible without a certain degree of creativity, unless complemented with qualitative studies to control for other intervening variables like gender. Relying on big data alone is not sufficient to ascertain, for instance, whether a laughing face was meant seriously or ironically, and therefore whether the person in question was laughing along with the joke or laughing at the person.

One common strategy to minimize bias and increase validity consists in cross-validation with a mixed method design. Several researchers from different disciplines reported seeking the confluence of evidence. Whereas a sociologist made efforts to compare the evidence of self-reflexive accounts of how people feel with the measuring of arousal by means of skin conductivity when using a smartphone, research teams composed of anthropologists and psychologists opted for a combination of long-term participant observation and experiments adapted to the local situations in order to obtain a variety of data on the socialization of emotions in different collectives. This approach enabled an expanded scope in considering mutual shapers and contextual factors, as well as more depth in the form of data intensity. While these two examples only point out the most innovative realizations of mixed method designs, data plurality is common in most ethnographic approaches. During field research, an interview can yield an audio recording, interview transcript, observation protocol, and entries in a field diary. Data plurality, then, includes not only speech data and written narratives, but also observations of facial expressions and body language. Although mixed methods approaches and source triangulation have been criticized for their failure to include full context information, the aim to strive for rich and comprehensive accounts should be acknowledged.

Moreover, when publishing results, scholars usually embed their data in theoretically founded arguments and commit themselves to theoretical candour. Together with fieldnote evidence and the elucidation of the ethnographer's path, this constitutes the three paths for ethnographic validity (Sanjek, 1990). Real difficulties are faced when research data are shared with other scholars. Secondary analysis as a form of recontextualization nevertheless requires additional information on the particular purpose for which the data were constructed so as to enable researchers to draw 'valid' conclusions. Categorical, dimensional or appraisal models to describe emotional behaviour can all be legitimate because validity also depends on the research questions and aims. Certain

questions – such as what might constitute context-sensitive validity criteria and whether recontextualization is only valid when preceded by reconstruction – will need be addressed in the future. Findings from the K-PLEX project indicate that little or no information on emotional aspects of the exchange between interviewer and interviewee is provided when data are shared (cf. Lehmann, Stodulka and Huber, 2018). Suggestions to place more emphasis on relational and situational aspects follow from this.

Patti Lather's (1993) fourfold approach to validity – subsuming situated, rhizomatic, ironic, and neo-pragmatic validity – can be useful in this regard. Situated validity refers to the situated, positioned, partial role of the researcher. It can be defined as a form of accountability as it calls for self-reflexivity and engagement, and connects epistemology and ethics. Several authors within the critical epistemologies' tradition have taken up these claims. Reflexive scholarship is then understood as engaging in discourse and encouraging research participants to articulate themselves. Validity can be formulated as a moral question of whose knowledge it is, how it is obtained, by whom, and for what purpose(s). Researchers and the research community cannot escape from their responsibility to make decisions on what is valuable and useful to study (Angen, 2000). To enhance the trustworthiness of research by reflexive positioning, four strategies have been suggested: identification with respondents; emotional enmeshment; personal associations; and auto-ethnography (Possick, 2009). Scholars in the field of emotion research, when performing the balancing act of caring about the research participants while remaining impartial to the data, have reported on the difficulties associated with such a dual strategy. Key challenges and loss of information due to methodological or ethical restrictions have mainly concerned the embodied and relational dimensions of affective and emotional practices through discourse, body language, gesture, and tone.

Rhizomatic validity that attempts to undermine authority, regularity, and common sense, and replace these with interactive co-construction and co-theorizing has been discussed above in relation to communicative validation and member checks. Ironic validity, the third component of Lather's approach, is associated with a transformative capacity whereby the author can position herself in a liberating way, emphasizing the ephemeral, rhetorical function of data. In emotion research, for example, scholars are well aware of the fact that data can have descriptive, performative, educative or still other functions. In particular, scholars who conducted research on emotion tracking acknowledged that individuals might 'attempt to play-act a little bit in line with these measurements, either to send the right messages to the machines or to kind of subvert or fake it, in order to navigate it'. Whether emotion tracking yields valid results depends on whether or not the person perceives the feedback to match one's own expectations. With self-tracking devices, the ways in which emotions are narrated continue to depend on habitus-based demands and social desirability. Depending on the object of research, the different dimensions of validity can intermingle in conflicting ways. Take, for example, the case of applied research in psychology that aims to design a self-tracking app to monitor and regulate one's emotions. At what point does a product designed for emotion recognition provide sufficiently valid data to be used for therapeutic purposes? If patient-centred validation is prioritized, then the assessment criteria might differ from those of the researchers. Patients might want to feel that they are engaged in social interaction when using the app, whereas the accuracy of the data might be the researchers' primary concern.

The fourth dimension of Lather's validity concept (neo-pragmatic validity) stresses the researchers' ability to tolerate the incommensurable, to deal with limits, paradoxes, and complexities. According to this principle, reporting ambiguities and contradictions is required if research is to be reported fairly. Going beyond the question of representation, validity can be defined as a 'space of constructed visibility of the practices of methodology' (Lather, 1993: 676). In the K-PLEX project, interviews with scholars in the field of emotion research highlight this challenge by disclosing how speech, texts, and images do not indicate emotions, but are rather proxy measures for emotions. To address emotions in their full complexity, researchers differentiated between enduring affective atmospheres and concrete emotional states. Tensions between purism and pluralism (Whittemore, 2001) can thus also be found in emotion analysis. Strategies to circumvent measurement biases that only consider acute intensities and no background emotions include: working in teams; comparing results with those of neighbouring disciplines; peer review; and combining deductive and inductive techniques. Thus, standards of judgement are as much representational as they are normative.

¹² These principles can be seen as equivalents to the four standards of evaluation, as there are interpretive, procedural, emancipatory, and postmodern.

Conclusion

Validity affects all stages of emotion research: perception; processing; evaluation; interpretation; and (self-) report. The kind of knowledge perceived as valid – whether introspective knowledge, empirical knowledge or practical knowledge – cannot be defined in general terms, but rather depends on the research purpose. Criticisms have been directed at the narrow focus on individual consciousness, experiences, and expressions. In order to increase the validity of research data, the embeddedness of individuals in social situations, discourses, and social structures should be taken into account. This kind of framework builds on long-established discussions about scale and the opposition of public and private fields. It reiterates the notion that common behaviour and official discourse in a group does not determine the behaviour and speech of individuals. Doing justice to the complexity of both the collective and the individual scales can help to prevent scholars in the field of emotion research from reproducing and validating cultural stereotypes. Data and especially ‘emotion data’ can be better understood, integrated, and shared when framed and contextualized as fluid, contextual, and relational. Institutionalizing doubt and remaining modest when seeking to verify correlations, always cognizant of the fact that ambiguity cannot be dissolved entirely, can guarantee a reasonable reporting of data and results. This article suggests that techniques for establishing or demonstrating validity might not always be feasible. Coping strategies can be found in altering techniques. Emphasizing the credibility of data analysis and presentation can compensate for impaired validity in the research design and data generation.

References

- Angen, M. J. (2000) Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue, *Qualitative Health Research*, 10(3), 378-395.
- Beatty, A. (2019) *Emotional Worlds: Beyond an Anthropology of Emotion*, Cambridge University Press, Cambridge.
- Creswell, J. W. (2013) *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 3rd edition, Sage, Los Angeles.
- Lather, P. (1993) Fertile obsession: Validity after poststructuralism, *The Sociological Quarterly*, 34(4), 673-693.
- LeCompte, M. D. and Goetz, J. P. (1982) Problems of reliability and validity in ethnographic research, *Review of Educational Research*, 52(1), 31-60.
- Lehmann, J. and Huber, E. (2018) Lost in datafication? A typology of (emotion) data contextualization, *Integrative Psychological and Behavioral Science*, accessed at: <http://dx.doi.org/10.1007/s12124-018-9470-6> (8 March 2019).
- Lehmann, J., Stodulka, T. and Huber, E. (2018) H2020 Project K-PLEX: WP4 Report on *Data, Knowledge Organisation and Epistemics*, accessed at: <https://hal.archives-ouvertes.fr/hal-01761214> (8 March 2019).
- Possick, C. (2009) Reflexive positioning in a politically sensitive situation: Dealing with the threats of researching the West Bank settler experience, *Qualitative Inquiry*, 15(5), 859-875.
- Sandelowski, M. (1993) Rigor or rigor mortis: The problem of rigor in qualitative research revisited, *Advances in Nursing Science*, 16(2), 1-8.
- Sanjek, R. (1990) On ethnographic validity, in R. Sanjek, ed., *Fieldnotes: The Makings of Anthropology*, Cornell University Press, Ithaca, NY, pp. 385-418.
- Whittemore, R., Chase, S. K. and Mandle, C. L. (2001) Validity in qualitative research, *Qualitative Health Research*, 11(4), 522-537.