

The experience of and reflection on triangulation and/or mixed methods, discussing a study on the ideal and reality, use and understanding of history textbooks

Bodo von Borries* – *University of Hamburg, Germany*

Abstract

This paper reflects on the achievements of triangulation but also on its inconsistencies and contradictions. It suggests that mixed methods can be confusing. Good triangulation should involve intensive reflection on a variety of materials and analyses, as well as a necessary complexity of questions and results. This implies a tolerant overlap between single-method and mixed-methods approaches. The paper illustrates these points by discussing an early example of mixed-methods research, a large project – Understanding of Textbooks, Use of Syllabuses and Processes of Reflection in History Lessons – carried out by the author in 2002. The aim of this study was to evaluate history textbooks and the ways in which historical sources and narratives were presented and used. It involved both a closed format, which produced quantitative data – a questionnaire, used to collect data on participants' knowledge, attitudes, emotions, and competences – and qualitative data collected from a subgroup of participants, who wrote short essays related to sections of the questionnaire. Some were interviewed after completing the questionnaire ('stimulated recall'). In this way, a type of two-step triangulation study emerged involving micro- and macro-analysis. The paper reflects on positive aspects of the study and also on its weakness. Participants' responses showed almost no correlation between their definition of the ideal textbook, which would deal with controversial issues, and the ways in which they used textbooks, with no recognition of the ambiguity.

Keywords: Germany; empirical studies; incoherence/inconsistency between beliefs and practice; mixed methods; triangulation

Character of the study

I do not remember when I heard or read the term 'triangulation' for the first time. Nevertheless, I know well the first moment I encountered a triangulated book. In 1971, I changed my own focus toward school and didactics and found a book on *Education and Societal Awareness* (Strzelewicz et al., 1966), an empirical *three-step study* combining *quantitative* and *qualitative* data, collected by questioning, group discussions and biographical interviews – before the term 'triangulation' appeared and spread outside the USA. I was very impressed – as were the public and the politicians (Sorge, 2016). I used the word to characterize a study done in 2002 and published with three younger colleagues, Claudia Fischer (sociology), Sibylla Leutner-Ramme (education) and Johannes Meyer-Hamme (history didactics) (Von Borries et al., 2005). There has been no equally important study in this area since, at least in the German

language, and we know from (rare) replication studies that the structures investigated are consistent and durable; thus, the results are still relevant after more than fifteen years. But what does triangulation mean? Since I have not studied theories and controversies in detail, but simply administered empirical studies, I will not distinguish between triangulation and mixed methods. In the first instance, I identify and equate both of them. Polemics aside, this appears to me as legitimate as any of the more elaborate taxonomies by experts on methodology.

For me, triangulation comprises a variety of structurally different approaches to the selected scientific topic, the researched problem, which can complement and thereby control one another. This methodological principle was known long before the term 'triangulation' was invented. I cannot imagine any well-done research in cultural anthropology – and micro-sociology of the poor – by fieldwork, without such mixtures of methods. To combine different approaches in an intelligent way has been the character of good scientific work from its very beginning, especially in the case of the empirical social sciences.

My report and reasoning are about the study, *Understanding of Textbooks, Use of Syllabuses and Processes of Reflection in History Lessons*, carried out in 2002 (Von Borries *et al.*, 2005; no English version). Due to a general scarcity of resources, the study suffered some disadvantages, especially the lack of a random sample. Instead we got a convenience sample of volunteers, in which self-selection easily causes a certain bias. There was not just one group of participants, but students of the sixth, ninth and twelfth grades (ages 11–12, 14–15, 17–18 respectively), their teachers and university students training to teach history. The respondents came from different parts of Germany and Austria, and from the German-speaking minority in Hungary. To distinguish all those statistical categories, as well as gender and types of schools, would have required a representative sample of enormous size (at least 5,000).

The overall count reached 1,361 ($N_{\text{MainGroup}} = 838$, $M_{\text{EssayGroup}} = 453$, $N_{\text{Teachers}} = 79$). This allowed us to experiment with various approaches. For example, the majority of students had to fill in closed-format items only, but for a minority ('third split'), a large part of the questionnaire was substituted by eight short essay questions, thereby combining quantitative and qualitative data. This proved to be of great benefit, as it mixed the methods. An even smaller sub-sample was asked to take part in interviews after the questioning. This thinking aloud (stimulated recall) produced two-step research as another form of triangulation, and allowed more detailed case studies than the short essays.

We chose to research all approaches to history textbooks. In a sort of opinion poll we asked for people's ideas on the ideal textbook, an evaluation of the textbook actually used, the perception of the everyday use of this textbook and the epistemological character of primary sources and historical depictions. All these items were given to learners and teachers (for sources and depictions, university students had to substitute for the teachers). Questions about concepts of religion, stereotypes of the Middle Ages, and evaluation of institutions and values were also included as background information. The same was true for the – very limited – testing of some basic knowledge about things such as the Ten Commandments and the sequence of eras in history.

The general idea was to combine questioning with an experiment in working with history textbooks or, more specifically, the critical comparison of three excerpts from textbooks on the same topic. Unfortunately, these were not given to the teachers, who got other controversial material. Thus, teacher training students had to substitute for the in-service teachers. The participants were asked to show how they understood

the information in the texts, the similarities, differences and (open) contradictions. This included not only reading ability, but also the competences of re-construction and de-construction. The test was no longer a mere opinion poll, but a measuring of historical competence in the full sense. In Germany, many theoretical models of 'historical competence' – others use the plural form – have been discussed. One of the most common was produced by a group of academics and best-practice teachers (Körber *et al.*, 2007) and is the only one offering a graduated scale and much empirical confirmation. (An older version was used in our 'triangulated' study.) The competence of 're-construction' means the ability for students to produce their own historical narrations from diverse material. 'De-construction' means the ability to analyse, examine and evaluate the previous historical narrations of others.

Cohort study and replication

The first example of triangulation is an easy one: the comparison of stereotypes about the Middle Ages (ranked on Likert scales with five steps) among students of different grades, including teacher training students. The differences between the groups in the same year (2002) are interesting in themselves (Von Borries, 2005: 49). Additionally, they can be taken as an approximation of the learning process between sixth grade and university. It is very unlikely that the ninth-graders of 2002, when they were sixth-graders in 1999, would have had very different ideas from those of the sixth-graders of 2002. Thus, the different cohorts of one year can be taken as a substitute for a longitudinal study, which would have needed far more resources. This is a not unusual – although a problematic – method and already forms a kind of triangulation.

Apparently, some stereotypes about the Middle Ages are accepted by many of the learners, especially the older ones (twelfth grade and university). The socio-historical concept of 'rule of the church and the kings over the peasants' and the politico-historical concept of 'quarrels between king and church' are the best examples. They are the most common from the outset, and are further strengthened in the process of education. The romantic interpretations of the epoch as 'adventurous' and 'glorious' – as well as the critical one ('dark and superstitious') – are weakly agreed by the younger participants (sixth grade), but clearly rejected by the eldest. They are partly extinguished or unlearned, although often not before university.

Eight years earlier, a representative sample (N = 2,007) of ninth-graders was asked identical questions (Von Borries, 1995: 57–60). The 2002 sample accepted all items slightly more readily than that of 1994. This can easily be explained by the fact that all participants in 2002 did the questionnaire willingly (self-selection) and therefore may have had a certain favourite bias. Generally, the structure and the priorities are identical. Of course, this is by no means an ideal case of scientific replication, but a first step toward that moderate type of triangulation.

Scales, means and correlations

Subsequently, we compared the conceptions of an 'ideal history textbook'. Of 14 items, 13 items could be combined into two sufficiently reliable scales (Cronbach's $\alpha > .70$). The naive and illusory conception of, or hope for, a textbook as an easy and pleasant representation of past reality ($\alpha = .79$) is at first highly present, but afterwards unlearned, although rather slowly, and to a neutral rating only. Simultaneously, the other more critical and more realistic concept of a textbook as a methodical and pluralistic encouragement towards historical thinking ($\alpha = .73$) is moderately accepted from

the outset and learnt even more. Among the twelfth-graders, it has already overtaken the naive version. But the real jump takes place between school and university (with the simultaneous selection of the cross-sample from all people to a tiny group of history specialists!). A realistic concept of ideal textbooks is not common among high-school students, not even in the highest grades (Von Borries, 2005: 63, 154).

Two other phenomena could be observed. Although we used a small and biased sample, the two sub-samples (the main group with closed-format questions only, and the essay group with combined closed-format and short-essay items) show nearly identical mean values. The other point is that the teacher training students hold faster to the methodical and pluralistic inclination towards historical thinking than the in-service teachers. The reason may be that they are younger and have heard more about history as a necessarily narrative and constructivist structure.

Parallel to the concept of the ideal textbook, we made the whole sample evaluate the textbook currently in use, the ways in which the textbook was used in lessons, and the concepts of primary sources and historical depictions (*ibid.*: 64–8, 68–71, 71–5). Thus, we could analyse the interrelations between these attitudes, convictions and habits, mostly by correlations of scales and comparison of mean values. The epistemological basis of history, the distinction between sources and historiography, was widely unclear, even among twelfth-graders. The articulations of teacher training students went in the theoretically correct direction, but not in a decisive manner.

Inconsistent articulations and non-compensating behaviour

The contradictions and inconsistencies between the articulations about attitudes and behaviour were evident and disturbing. In the triple estimation of the ideal textbook, the existing textbook and actual use of the textbook, many questions address very similar, or even identical, categories. But the students' answers show zero correlation in nearly all cases; a common variance of more than 2.6 per cent occurs only once. The situation is even worse in the case of the teachers: they observe sizeable differences between the ideal and the actual textbook. The comparison of the ideal textbook and actual tasks given to the students for textbook work in the teachers' own lessons is even more striking (*ibid.*: 115–16).

In the ideal sphere, teachers prefer the progressive tendency of a controversial and open character, which, in their perception, is not achieved at all in the existing textbooks (*ibid.*: 115). Therefore, one would expect that the teachers explicitly try to compensate for the deficits of the books by the methods they adopt and the tasks they set, or give the learners much freedom and autonomy, but this is by no means the case.

The teachers abhor any unambiguity in history, but unexpectedly trust in the fundamental facts of history. They have not noticed a connection between both categories (zero correlation). In the more extreme case of the real past, a positivistic misunderstanding, there is neutral acceptance and a very sizeable correlation (48 per cent common variance) between ideal and own practice. In terms of easiness and fun or fascination, the teachers perceive very clearly that the existing textbooks are far behind the ideal (*ibid.*). But again, no attempt at compensation is made at all, in favour of catering for the students' enjoyment and giving them the necessary space for their individual development.

Experimental comparison of excerpts from textbooks

We made the participants read, examine and compare three excerpts on Boniface, the 'Apostle of the Germans', from three textbooks of quite different types. The Anglo-Saxon missionary (c.673–754/5) cut down a Donar's oak in Hesse in 722/3 and baptized many 'heathen' people. After having organized some bishoprics and abbeys in the East of Francia, he went to Frisia, preaching as an old man, and was killed there by natives in 754/5.

Afterwards, their understanding of the texts (that is, the correct reception of the information), their emotions while reading and working, while experimentally taking the perspective of a contemporary Frisian youngster in 754/5, and their own sense-making (relating history to today) were assessed, mostly by five-step Likert scales. We obtained a series of tolerably consistent – although not ideally reliable – scales (*ibid.*: 77, 87, 91, 94).

The sixth-graders hardly exceed chance (the probability of mere guessing) when acquiring the information. Ninth-graders are a bit stronger, and twelfth-graders still exhibit a lot of problems (not much more than 50 per cent correct solutions). Even the university students refuse to answer or fail in a quarter of the questions. The selected textbooks were intended for sixth or seventh grades. Although it is not the first time that such results have been found (Von Borries, 1992, 1995), neither ministries nor publishers show any interest in these disappointing empirical results. Students are fundamentally overtaxed by their textbooks. Two of the excerpts contained a significant contradiction of simple facts. Nevertheless, students did not notice this inexcusable mistake. Their misunderstanding of the texts is profound. From the items on their own estimation of the existing textbooks, we know that they describe them as quite easy or at least easy enough. They live in serious self-delusion.

Spontaneous emotions, experimental empathy and personal consequence have means very near to the neutral point. The life of Boniface – his mission and his killing – do not really delight, offend, or consternate young people today; they seem to perceive them as rather indifferent and remote. Generally, our experiment on the reception of the textbooks about Boniface was not very successful, mainly because it did not really interest or engage the students. With other (more thrilling) dilemmas (witch hunts, war of aggression, massacre during a crusade, forced marriage), we obtained much better results (Von Borries, 1992, 1995, 1999). Perhaps such categories are seldom or never mentioned in their history lessons. The development between sixth-graders and university students (who are young near-experts) is small, but characteristic. All feelings are reduced over the course of time. At the outset, the positive emotions are a bit stronger but in the end it is the negative ones that prevail.

In the case of a contemporary's hypothetical decision in 754/5 between slaying the missionary Boniface or getting baptized by him (as chosen in 723/4 by the Frisians), all responses indicate opposition to the killing; that includes a clearly counterfactual position and argument. The contemporary acceptance of baptism (as had been the Hessians' choice in 723/4) shrinks with age and expertise (again, more counterfactual than historical). Both phenomena of making a moral decision, instead of taking historical perspectives, have been found in earlier studies mentioned above (Von Borries, 1992, 1995, 1999).

We are not really surprised that the readiness of linking past and present, of orientation from history in different logical patterns (Rüsen, 1994: 37–41, 85–90, 150–5, 231–3), is reduced among the older learners in the case of Boniface. Neither a traditional-affirmative nor a critical-distanced relationship is accepted by the older

students. Genetic sense-making could only be identified in a third (weak) factor. (Perhaps we did not find adequate items to assess it.)

Quantification of qualitative data

We also tried to test our learners' competence of reconstructing and deconstructing with closed-format item groups about the three Boniface textbook versions. In this purpose we failed completely, and with absurd results (ibid.: 95–104). Therefore, we had to take an additional step: the combination of quantitative and qualitative data. We changed the item format for the last third of the total sample and demanded short essays. Thereby we obtained completely different findings. Of course, we had to first categorize and quantitatively rate the texts – again on five steps from '1' to '5' – (Von Borries, 2005: 132, 145, 150, 152; 2012: 61). The learners had not really been able and/or willing to read accurately and to distinguish precisely between our different questions. They often wrote 'I have answered this already above!' or something similar. Therefore we had to aggregate the eight short essays of every individual first, and then relate the whole (comprehensive) text to any of our specialized questions. Unfortunately, we had no resources for a double rating (and thereby a control of inter-rater reliability). Instead of that, we encoded the material once (temporarily doing away with gender and grades to assure a *blind* method) and discussed all dubious cases (about 5 per cent of the sample) intensively.

In a factor analysis, the quantified qualitative data yielded four factors; that is, four uncorrelated (orthogonal) dimensions: deconstruction, reconstruction, textbook preference and type of sense-making. Apparently, the qualification of reconstruction ($M_{\text{reconstr.}} = 1.42$) is higher than that of deconstruction ($M_{\text{deconstr.}} = 1.09$) from the outset, and it also grows much earlier and faster. Among the twelfth-graders, the difference is considerable ($M_{\text{reconstr.}} = 2.36$, $M_{\text{deconstr.}} = 1.67$). Nevertheless, in the case of the university students it has nearly vanished ($M_{\text{reconstr.}} = 2.60$, $M_{\text{deconstr.}} = 2.46$). This is a very important point and has never been measured before (as far as we know). Two different causes may be considered. Deconstruction could be (epistemo-)logically more difficult and complicated. Even more likely, students are only infrequently taught to use it and with little emphasis, except at university level.

Even more strikingly, we could measure three of Rösen's (1994: 37–41, 85–90, 150–5, 231–3) four patterns of sense-making. The exemplary mode could not be isolated precisely, perhaps because it is the most common and frequently encountered one. From the outset, traditional argumentation ($M_{\text{sixth-graders}} = 1.53$) is by far stronger than genetic argumentation ($M_{\text{sixth-graders}} = 1.03$), and critical sense-making ($M_{\text{sixth-graders}} = 1.30$) is in the middle. This structure remains stable, while all three types grow constantly, but at different rates. Even among university students, only traditional sense-making ($M_{\text{univ. students}} = 3.01$) reaches a neutral level; the critical type ($M_{\text{univ. students}} = 2.12$) and especially the genetic one ($M_{\text{univ. students}} = 1.25$) are simply refuted, even by the university students (as near-experts). It seems that school and university teachers are very cautious and normally avoid thinking about the relationship between past and present. Thus they do not follow the 'new' – now fifty years old! – concepts of history didactics at all.

Two-step research and case studies

In the essay group, quantitative data are available for all participants about most questions, as well as short essays on some questions. Thus, every individual can be sorted or ranked in his or her peer group and characterized as individuals at the same

time, as case studies demonstrate (Von Borries, 2005: 158–74; 2006: 120–2 and CD-ROM 1–6). Of course, these may be managed in more detail and more fruitfully if we can interpret thinking aloud data or – even better – biographical interviews (Von Borries, 2005: 174–80, 181–216). In our study, we experimented with both possibilities, and we succeeded to a certain degree. Only two participants, both of them intellectually far above the average for their peer groups, will be compared here. Comparisons with others of the same age groups show that they offer longer explanations, greater differentiation and clearer argumentation; the higher quantitative scoring reflects this.

The difference between both groups is obvious, not only in length. Jana (sixth grade) – although proudly describing the questions as easy – omits autonomy and empathy and misunderstands perspective, commonalities and differences. Nevertheless, she has a clear idea about the legitimization of the topic ‘Boniface’ – eagerness and explanation of the present via tradition, while her sense-making is obviously an exemplary one. She writes about her emotions and not about the layout of the books, as most of the other students of all grades do. Her emotions are apparently steered by moral principles (not by historical reflections), which do not allow her to accept slaying, idolatry or execration, but claim miraculous heroism and Christian mission. Jana prefers the old-fashioned and pictureless compendium. In this, she chooses what we have identified as the easiest textbook. The majority within her peer group show a predilection for the other two textbooks.

Marianne (twelfth grade) is one of the very rare participants who detects a severe contradiction between two books in the fundamental description of the existing primary sources. This is achieved by only a tiny minority, even among the university students, and is no marginal point, but central to methodical awareness. When asked for her emotions, she articulates almost expertly the didactical quality of the books, thus anticipating the last question (and making a false guess about the dating of the textbooks). Even in twelfth grade, and by an excellent student, the tendency of this statement is completely opposed to that of modern academic theory about the importance of affective, empathetic, emotional understanding.

Marianne sees the legitimation of the chapter in the textbooks correctly in the tradition, or – more precisely – in the supposed and desired identities of readers and learners. Her own self-evident, unquestioned identification, shown by writing ‘our’ three times (for history, country and faith), is very important. She sustains the old downgrading dichotomy between ‘us’, the good Christians, and ‘them’, the savage heathens. She continues this tendency when answering the question of relevance for today with an answer displaying merely traditional sense-making: she assures us that Boniface has caused the uniformity of today’s Germany, when an impartial observer would be hard pressed to notice any such uniformity. History can and will be (mis)used to solidify evidently wrong interpretations of present situations, even by high-school learners.

Tentative generalizations

The attempt to measure students’ historical competence in 2002 generated frustrating results. Our study of 2002 is far from ideal or perfect, but – in respect to the suspicion that triangulation is much more a desideratum in handbooks of methods than a practice in everyday research – it is worth presenting it here and reflecting on it as an effective example. Indeed, it fulfils all four classical types of triangulation (Denzin, 1970). *Data* are triangulated by asking school students of different grades, university students and in-service teachers, additionally in different countries and at different times. *Researchers*

are triangulated by the cooperation with three younger colleagues from different disciplines. *Theories* are triangulated already by the simple fact that history didactics work in a multi-, inter- and transdisciplinary field between pedagogy, psychology, sociology and history. *Methods* are triangulated by closed-format items, short essays and transcribed interviews, by statistical analysis, quantification of qualitative material and hermeneutic interpretation. The outcomes, especially the epistemological deficits, were equally disappointing when we tested the competence of ninth-graders in the project HiTCH (Historical Thinking – Competencies in History, 2012–15) more than ten years later. This time, we used quantitative formats of psychometrics only, meaning without triangulation, but employing other methods of validation (Trautwein, 2017).

Any discussion about equal and interdependent ranks of quantitative and qualitative studies is superfluous. The ‘silver bullet’ may be a combination of both approaches, for example, in the form of two-step research, beginning with a representative survey and going on with interviews or experiments with some individuals from the average and from different outlier groups. The distinction between quantitative and qualitative is not absolute, either. A lot of qualitative data (such as our short essays) can be transformed into quantitative data by encoding or rating. In other cases (for example, video or whole lessons), encoding cannot easily reach low-budget operationalization and quantification.

Indeed, the type and peculiarity of triangulation seem different while using qualitative and quantitative – or quantitatively rated – material. In this second case, after the construction of scales or factors, correlations and cluster analysis may be the most productive techniques for finding and controlling interrelations (even causalities, where additional arguments exist). Apparently, the opposite is as important. Often there is no significant or relevant correlation, meaning there is no connection or no coherency. Although falsification (in the meaning of contradiction, not of forgery!) may be frustrating, it surely is as fruitful as verification (which – according to Popper – is impossible, and therefore should be called provisional validation).

Triangulation may be more useful for showing the over-complexity of reality and the incoherency – also inconsistency – of human behaviour between emotion and cognition, knowledge and acting, norms and wishes, and so on (complication) than for confirmation of rather simple connections (validation). It may help to improve theory models. Another important purpose is connecting surveys and case studies (combination of micro- and macro-level). Triangulation can remain constantly aware of the following: results (articulations, reactions, behaviour) depend very much on the concrete situations and circumstances, and therefore contexts have to be described, included and considered very carefully.

If we look at qualitative material, we might have used it ourselves or discovered it (found others’ pre-existing articulations). The latter has some advantages of saving money, time and labour, and avoiding method effects and going back to past epochs. Favourite examples are my own studies about historical socialization (Von Borries, 1996a: 79–103, 17–37), parental punishment (Von Borries, 1996b), students’ school experiences (Von Borries, unpublished), and strategies of marriages (Von Borries, 2003: 257–98) using hundreds of German-language autobiographies of the eighteenth to twentieth centuries. I cannot explain the adequate techniques or strategies of an elaborate triangulation here. Many classical models of content analysis and interpretation via systematic and complete double encoding (to ensure inter-rater reliability) are impossible strategies in this case, because they require too much money, work and time in the face of a sample of about three hundred autobiographies – from the eighteenth century alone – with an average of about three hundred pages

each. But, surely, enlightened hermeneutics includes mechanisms of control and reflection anyway.

Notes on the contributor

From 1976 to 2008, **Bodo von Borries** was Professor of Didactics of History at Hamburg University (Germany), researching concepts of historical learning, analysis of history textbooks, the production of alternative teaching material (non-European history, women's history, history of childhood, environmental history) and empirical research of historical consciousness, including representative cross-cultural studies in East and West Germany and in Eastern and Western Europe after 1989–91.

References

- Denzin, N.K. (1970) *The Research Act: A theoretical introduction to sociological methods*. Chicago: Aldine.
- Körper, A., Schreiber, W. and Schöner, A. (eds) (2007) *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: Ars Una.
- Rüsen, J. (1994) *Historisches Lernen: Grundlagen und Paradigmen*. Köln: Böhlau.
- Sorge, V. (2016) *Die Göttinger Studie und ihre Auswirkung auf den Bildungsbegriff in Westdeutschland*. München: GRIN.
- Strzelewiec, W., Raapke, H.-D. and Schulenburg, W. (1966) *Bildung und gesellschaftliches Bewußtsein: Eine mehrstufige soziologische Untersuchung in Westdeutschland*. Stuttgart: Enke.
- Trautwein, U., Bertram, C., Von Borries, B., Brauch, N., Hirsch, M., Klausmeier, K., Körper, A., Kühberger, C., Meyer-Hamme, J., Merkt, M., Neureiter, H., Schwan, S., Schreiber, W., Wagner, W., Waldis, M., Werner, M., Ziegler, B. and Zuckowski, A. (2017) *Kompetenzen historischen Denkens erfassen: Konzeption, Operationalisierung und Befunde des Projekts "Historical Thinking – Competencies in History" (HiTCH)*. Münster: Waxmann.
- Von Borries, B. (1992) *Kindlich-jugendliche Geschichtsverarbeitung in West- und Ostdeutschland 1990*. Pfaffenweiler: Centaurus.
- Von Borries, B. (1995) *Das Geschichtsbewußtsein Jugendlicher: Eine repräsentative Untersuchung über Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen von Schülerinnen in Ost- und Westdeutschland*. Weinheim: Juventa.
- Von Borries, B. (1996a) *Imaginierte Geschichte: Die biografische Bedeutung historischer Fiktionen und Phantasien*. Köln: Böhlau.
- Von Borries, B. (1996b) *Vom "Gewaltexzeß" zum "Gewissensbiß"? Autobiografische Zeugnisse zu Formen und Wandlungen elterlicher Strafpraxis im 18. Jahrhundert*. Tübingen: Edition Diskord.
- Von Borries, B. (1999) *Jugend und Geschichte: Ein europäischer Kulturvergleich aus deutscher Sicht*. Opladen: Leske und Budrich.
- Von Borries, B. (2003) *Wendepunkte der Frauengeschichte II: Über Muttergöttinnen, Männeransprüche und Mädchenkindheiten: Modelle und Materialien zum Ausprobieren und Bessermachen*. Herbolzheim: Centaurus.
- Von Borries, B. (2006) 'Fähigkeiten zur De-Konstruktion von Geschichts-Schulbüchern: Empirische Befunde (2002) und fachdidaktische Vorschläge (2004)'. In Schreiber, W. and Mebus, S. (eds) *Durchblicken: Dekonstruktion von Schulbüchern*. 2nd ed. Neuried: Ars Una, 125–32, with documentation on CD-ROM.
- Von Borries, B. (2012) 'Erwartungen an, Erfahrungen mit und Wirkungen von Geschichtsschulbüchern – empirische Befunde'. In Doll, J., Frank, K., Fickermann, D. and Schwippert, K. (eds) *Schulbücher im Fokus: Nutzungen, Wirkungen und Evaluation*. Münster: Waxmann, 43–65.
- Von Borries, B., Fischer, C., Leutner-Ramme, S. and Meyer-Hamme, J. (2005) *Schulbuchverständnis, Richtlinienbenutzung und Reflexionsprozesse im Geschichtsunterricht: Eine qualitativ-quantitative Schüler- und Lehrerbefragung im Deutschsprachigen Bildungswesen 2002*. Bayerische Studien zur Geschichtsdidaktik, Vol. 9. Neuried: Ars Una.