

On academic reading: citation patterns and beyond

Anton Oleinik^{1,2}  · Svetlana Kirdina-Chandler³ ·
Irina Popova^{4,5} · Tatyana Shatalova⁶

Received: 27 April 2017 / Published online: 24 July 2017
© Akadémiai Kiadó, Budapest, Hungary 2017

Abstract The article discusses the process of textually mediated communication in science and proposes an approach that complements citation analysis. Namely, it addresses the question of how the author's text is read by the reader and whether the reader interprets the text in the same manner as the author. Fifty-seven scholarly contributions (articles, book chapters and book reviews), written by three social scientists, were content analyzed with the help of the *QDA Miner* and *WordStat* computer programs. The outcomes of the qualitative coding were compared with the outcomes of the analysis of word co-occurrences and the outcomes of the analysis on the basis of a dictionary based on substitution. Our findings suggest that texts have plural interpretations. Depending on the reading context, either the author's or the reader's perspective prevails. Also, both the author and the reader may read the text in a either deep or perfunctory manner. Deep reading necessitates spending significant time and cognitive resources.

Keywords Academic reading · Comprehension · Interpretation · Citation analysis · Content analysis

✉ Anton Oleinik
aoleynik@mun.ca

¹ Department of Sociology, Memorial University of Newfoundland, Elizabeth Avenue, St. John's, NL A1C 5S7, Canada

² Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russia

³ Institute of Economics of the Russian Academy of Sciences, Nakhimovskiy Avenue, 32, Moscow, Russia 117218

⁴ Institute of Sociology of the Russian Academy of Sciences, Krzhizhanovskogo Street, 24/35, Build. 5, Moscow, Russia 117218

⁵ National Research University – the Higher School of Economics, Moscow, Russia

⁶ XXI Century Integration International Secondary School, 16 Marshala Katukova Street building 3, Moscow, Russia 123592

Introduction

Despite the claims made by some specialists in communication studies that “communication of science is textual” (Budd 2001, p. 314), studies of depersonalized communication in academia leave several gaps in our knowledge of this topic. First, academic reading and writing is rarely placed in the context of social action. “Action is ‘social’ insofar as its subjective meaning takes account of the behavior of others and is thereby oriented in its course” (Weber 1968, p. 4). The writer and the reader interact, i.e. influence one another (Yore et al. 2002, p. 675).

Second, scholars of science communication pay relatively less attention to communication to fellow scientists than to communication to other groups: the general public, students, practitioners, and government officials. For instance, a search carried out in the *Web of Science* database produced 552 texts on the topic of adult reading, 374 texts on the topic of school or student reading, 133 texts on the topic of children reading and 21 texts only on the topic of academic or scholarly reading.¹

Third, when peer-to-peer text-based communication in science is actually studied, citation analysis prevails. Citation analysis focuses on quantitative aspects of reading. It is based on the assumption that the reader’s understanding of the author’s ideas can be operationalized by reading a citation.

In this article, sociological approaches are applied to the study of text-based communication in academia. We consider scholarly texts as necessary preconditions for and outcomes of depersonalized interactions. This approach is in keeping with viewing texts “as the interactions of people who are members of specific communities” (Hyland 2004, p. 132) and “writing as knowledge building rather than knowledge telling” (Yore et al. 2004, p. 353). We study the reading of scholarly texts with the help of an original strategy for mixing qualitative and quantitative data, which complements citation analysis.

This article focuses on discussing the problematic nature of scientific communication in the form of reading, writing and citing. When communicating at a distance, scholars do not have several of the supports that are available in face-to-face interactions. For instance, when assessing the validity and reliability of one’s work, they do not necessarily know that individual’s personal reputation. As a result, their judgment is based mainly on what is provided in the text. The relevant personal characteristics of the author—his/her intentions, interests, membership in various networks, honesty and integrity—remain hidden behind the text.

A specific research question derives from the defined general framework of inquiry: how is the author’s text read by the reader? This question can be addressed from the author’s point of view. Then it reads as follows: how closely does the reader comprehend the author’s thought? The reader’s task consists in receiving the author’s message with minimal losses of information and in correctly “deciphering” it. If one adapts the reader’s perspective, the emphasis in the research question shifts accordingly: how relevant is the information provided by the author to the reader’s situation? Finally, viewed from an outside perspective—from the perspective of an impartial observer—the question has to be reformulated once again. Now the emphasis is placed on new meaning emerging in the process of reading: does the process of reading produce new ideas in addition to those of the author and the reader?

¹ The search was conducted on April 25, 2017 using the search terms TOPIC: (“adult reading”), TOPIC: (“school reading”) etc.

This article has three sections, in addition to the introduction and conclusion. In “Communicating at a distance in science” section provides a view of text-based communication in science with the help of a review of the relevant literature. In “Case study of reading” section discusses the sources of the primary data. Four social scientists read and content-analyzed each others’ works. The techniques of content-analysis served to quantify the outcomes of the reading and to provide a qualitative description. The four social scientists’ readings are compared and discussed in “The battle between author and reader” section. Namely, the author’s reading of his/her own texts is compared to the readers’ interpretations. It is shown that, in some reading contexts, the author’s view prevails whereas, in other cases, the reader has the upper hand in attributing meaning to texts.

Communicating at a distance in science

Textually mediated social action in science can be represented as a set of transactions between three types of actors (Fig. 1). The Author of a contribution (an article, a letter, a book, and so forth) addresses his/her message to a personally unknown Reader. When writing it, the Author refers to the body of the existing literature: he/she cites the author of some other work (the “Cited Authors”). By doing so, the Author also plays the role of a Reader with respect to the cited source. The Reader eventually turns into the Author of the next generation of contributions. As Latour (1987, p. 38) aptly put it, “to survive or to be turned into fact, a statement needs the next generation of papers (citing it)”. Consequently, the transaction between the Author and the Reader represents an element of an endless chain of scientific communication.

To provide more nuances for the scheme, the singular noun “Reader” needs to be changed to the plural form. As a matter of fact, the Author addresses a potentially unlimited number of personally unknown Readers (Reader_i ... Reader_k). Each of them reads the text in a different manner in keeping with his/her particular situation: interests, background knowledge, cognitive capacities, etc. (White 2011, p. 3348).

Nevertheless, these apparently trivial statements help to better specify the line of subsequent reasoning. Reading is an integral element of depersonalized communication in

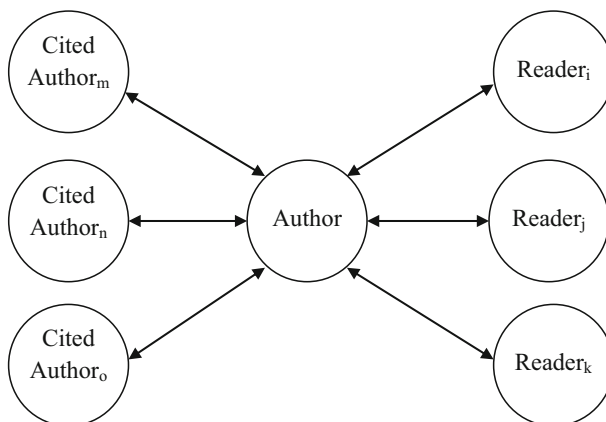


Fig. 1 Text-based scientific communication: basic scheme

science. And by comparing the Author's message with the Readers' interpretation, one gains a better understanding of the entire chain of scientific communication at a distance.

Figure 1 also serves to better position the approach proposed in the present article with regards to the existing literature on science communication. A large body of research deals with text-based communication to the general public (Hultberg 1997, p. 194; Littmann 2005; Brossard and Shanahan 2006; Fischhoff 2013). These studies address the question of how good scientists are at conveying their knowledge to non-specialists and at educating citizens. Science communication scholars also pay significant attention to the issue of communication to students (Abbott 2013), especially to students for whom English is a second language (Weigle et al. 2013). We propose to emphasize communication of scientists to fellow scientists, namely peer-to-peer communication in science. When doing so, we go beyond citation analysis (to what Cited Authors the Author refers in his/her contribution). Namely, we discuss how deeply the Author's contribution is read. What ideas do the Readers identify in the Author's contribution before they eventually cite it in the next generation of papers?

Reading varies in depth. A perfunctory reading differs in its intensity from a deep reading of the same text (Gryaznova and Ratz 2008). Perfunctory reading involves "screening" the text (Wilson and Tenopir 2008, p. 1395). The Reader looks for relevant information consulting the table of contents, title, and abstract. When the Reader goes to the body of the article, he/she interacts with the Author in a one-way manner. The Reader simply selects those bits of the Author's message that seem to be the most relevant for his/her current situation. The Author sends a message, the Reader receives it imperfectly (highly selectively). If the Reader's take on the text diverges from the Author's in the case of a perfunctory reading, this results from the former's lack of attention.

Deep reading requires the Reader to invest more cognitive resources and time, which allows him/her to "delve" into the text instead of remaining on its surface. If the Reader intends to use a particular text in his/her own work (to cite it, for instance), he/she usually reads it two or more times (Wilson and Tenopir 2008, p. 1403) and takes notes. Deep reading also involves a two-way format of communication between the Author and the Reader. The Reader considers the Author's arguments in the light of his/her situation. The Author's intentions do not necessarily coincide with the Reader's, which attaches a plurality of meanings to the same text.

In this sense, a deep reading is related to Bakhtin's notion of multi-voicedness (Bakhtin 1979, p. 41). The Author's voice represents one voice among many. He/she simply has a very limited control over what the Readers can make of it. Abbott (2013, p. 193) calls this approach to reading "active" and opposes it to "passive" reading. A "passive reader" reproduces the Author's key ideas through quotation. An "active reader" develops his/her own ideas in response to the Author's train of thought.

The difference between the one- and two-way communication of the Author and the Reader can be better understood with the help of the categories of comprehension and interpretation. Comprehension involves using the Author's message as a point of reference: the Reader comprehends a text correctly if he/she discovers the meanings attributed by the Author. The Reader has many more degrees of freedom when interpreting a text. Namely, he/she is free to attribute new meanings to it (Norris and Philips 1994, p. 402). Defined in this way, comprehension involves the prevalence of the Author's point of view, whereas interpretation makes the Readers' perspectives more important.

Weigle and co-authors operationalize the concepts of comprehension and interpretation and apply them to the study of academic reading. When analyzing students' reading skills, they differentiate a "text model of comprehension" and a "situation model of

interpretation” (Weigle et al. 2013, p. 29). The Readers—students for whom English is a second language in this case—use the former if they prioritize discovering the propositions within the text and their interrelationships. The latter model refers to the Readers’ attempts to get an individualized take on the text interpreting its key propositions in light of their own experience, previous knowledge and particular situation.

Some parallels can be drawn between the concepts of comprehension and interpretation, on the one hand, and two major epistemological positions, positivism and interpretivism, on the other. A divide between objectivists/positivists and constructivists/interpretivists characterizes the literature on science communication (Budd 2001, pp. 309–312; Stanovich 2003, pp. 107–109). With respect to the study of reading, the objectivist stance is based on the assumption that the truth exists objectively and independently of the observer (e.g. the Reader). The constructivist stance highlights subjective moments in the reading process: the truth (a truth, as a matter of fact) should make sense to the observer and correspond to his/her previous beliefs and knowledge. Thus, the objectivist truth is merely understood, whereas a constructivist truth requires interpretation.

In this study, the objectivist comprehension is operationalized through the assumption that the text as a particular constellation of words conveys a limited range of meanings. The Sapir-Whorf hypothesis states that language shapes its speakers’ view of reality (Whorf 1956; Sapir 1961). We operationalize the constructivist interpretation through the assumption that, in addition to the meanings embedded in the text, the Reader finds new ones. And these new meanings are specific to the Reader and his/her particular situation. Section “[Case study of reading](#)” will provide more detail about confronting the objectivist comprehension and the constructivist interpretation in empirical research on reading.

Speaking more specifically about the Reader’s stance toward the text, he/she adapts either a cooperative or a critical, adversarial approach. These two approaches could eventually complement one another, as in the case of *critique désintéressée* in the Republic of Letters. The requirement to judge “without taking sides” (Goldgar 1995, p. 113) necessitates both cooperation in the advancement of knowledge and confrontation of arguments. However, analytical as well as practical reasons dictate that conflict and cooperation in reading are considered separately. As opposed to an ideal citizen of the Republic of Letters, the actual Reader most often either cooperates with the Author or criticizes him/her (or simply disregards his/her arguments while being satisfied with a perfunctory reading).

By choosing a cooperative approach toward the text, the Reader further develops the Author’s line of reasoning. A cooperative reading “is oriented towards further use: it is highly selective, as well as highly interested in the direct application of earned knowledge in the reader’s own work” (Hirschauer 2010, p. 77). Namely, the Reader finds new applications for the Author’s argument or builds his/her own theory based on it.

Cooperation between the Author and the Reader requires a high level of generalized trust: in most cases, they do not know one another personally. The paradigmatic sciences—the disciplines in which scholars agree on basic assumptions—are characterized by a more pronounced tendency toward cooperation. A comparison of book reviews, on the one hand, in the arts and social sciences, and, on the other hand, in the natural sciences, suggests that praise prevails over criticism in the latter case (Hartley 2006, p. 1196).

A heavy reliance on trust in reading at the expense of criticism might lay, however, at the origin of some negative tendencies. For example, the Readers often forget to indicate page numbers in their references to particular texts. This forgetfulness is tolerated as long as the Reader—the Author in the making—appears trustworthy. The omission of page numbers in bibliographical references, initially common in the natural sciences, becomes

more and more widespread in the arts and social sciences as well. A discovery of even minor mismatches between what the Author really says and what the Reader attributes to him/her justifies a call for being as specific as possible when referring to a particular text (Henige 2006).

The choice of a critical, adversarial approach toward the text means that the Reader challenges the Author's assertions. The Reader does not take anything said by the Author for granted, questioning and criticizing everything instead (Abbott 2013, p. 194). When writing a text, the Author's major concern then consists in anticipating "possible negative reactions (of the Reader) to his or her persuasive goals" (Hyland 2004, p. 13; see also Latour 1987, p. 46). The Author, who correctly predicts and addresses eventual criticisms, has more chances of conveying his/her message. In other cases, the Reader either interprets the Author's message as he/she sees fit or simply ignores it.

The prevalence of criticism in scientific communication at a distance undermines any eventual cooperation between the Author and the Reader. A skeptical Reader (Spektor-Levy et al. 2009, p. 876) does not have much sympathy for the Author and his/her intents. Individuals in a common quest for the advancement of knowledge change into adversaries, if not enemies. The adversarial character of the interactions between the Author and the Reader suggests parallels with a judicial trial or even a battlefield (Latour 1987, pp. 58, 172; Hirschauer 2010, p. 78).

The practical application of the taxonomies of reading outlined above (deep versus perfunctory, cooperative versus critical) requires solving a number of methodological problems. One such problem refers to the issue of operationalization. How can the depth of reading or the intensity of criticism be measured? For instance, Bourdieu argues that the degree of hostility tends to be very high in scientific interactions. However, it rarely takes explicit forms (Bourdieu 1984, p. 39).

Citation patterns are commonly used as a proxy for both the depth of reading and the intensity of criticism. The analysis of citations serves to produce a quantitative measure of effectiveness of reading. The number of references to an Author's work is presumably indicative of the number of his/her Readers. By looking at the format of particular citations, one can deduce whether the Reader (Citing Author) adapts a critical or cooperative stance toward the Cited Author (Harwood 2009). Semantic analysis serves to further substantiate assumptions about the Reader's stance (Hyland 2004).

The Reader's cognitive resources and time are limited, which transforms into a particularly important constraint at the time of the proliferation of scientific publications. "Attention rather than information (becomes) the scarce resource" (Simon 1978, p. 13). Being unable to read everything relevant to his/her field of studies, the Reader attempts to optimize the use of his/her increasingly scarce attention. Relevance theory refers to the rule of best possible balance of effort against effect in this regard (White 2011).

How a text is read depends on its genre. Each genre of textually mediated communication—scholarly article, scientific letter, book, book review, abstract—implies particular requirements with respect to the structure and content of the text (Hyland 2004). Differences in the structure of scholarly texts either facilitate or complicate their reading. "Scientific articles, because they follow familiar formats and report on common procedures, are typically shorter than articles that lack these things" (Hartley et al. 2004, p. 189). Consequently, the reading of well-organized and structured articles requires relatively less effort.

As opposed to an article, a book (monograph) has a less rigid structure. The Author of a book discusses several ideas instead of just one (cf. the rule "one article—one idea"), develops his/her arguments in more detail and is free to explore the "side-paths" (ideas

that are only indirectly relevant to the main line of reasoning). As a result, the book calls for a deeper reading. Social scientists communicate mainly through writing and reading books. According to a study, monographs amount to 52% of cited references in the social sciences, as opposed to just 7% in medicine (Wilson and Tenopir 2008, p. 1398; see also Volentine and Tenopir 2013, p. 435).

Books can be further classified in two categories, specialized and overviews. The Reader of the former category is a peer specialist; the general public constitutes the readership of the latter category. Since reading specialized books requires concentration, the investment of a significant amount of time and substantial background knowledge, it appears particularly problematic. Publishers, including the university presses, increasingly favor overviews and essay-type books that can be rather easily “digested” by the Reader (Auerbach 2006).

The study of citation patterns does not cover all aspects of the relationship between the Author and the Reader. The underlying assumption involves attributing a single idea to a text. Then, the number of references to a text is, arguably, indicative of how many readers received this unique Author’s message. Plurality of citation functions undermines the belief that the Cited Author’s work was necessarily read, let alone interpreted in keeping with to his/her intentions (Harwood 2009). For instance, competence, tying or signposting citations suggest just the opposite: the Reader has barely familiarized him/herself with the text. Furthermore, a book or an essay usually contains more than just one idea. Thus, a reference to a book or an essay does not indicate which of the Author’s ideas caught the Reader’s attention.

The methodology of content-analysis helps to bridge the gaps left by the study of citation patterns. Qualitative content-analysis (manual coding) is intended to identify text fragments that correspond to the Author’s and Reader’s ideas and key concepts. There is no restriction as to their number. Quantitative content-analysis (co-occurrence of words) greatly facilitates the semantic analysis of sentences. The use of a dictionary based on substitution—a hybrid form of the qualitative and quantitative types of content analysis—paves the way to conducting qualitative content analysis in an automated regime. No human input is required after specifying the words and sentences that refer to each of the qualitative codes (the Author’s and Reader’s ideas and key concepts). Finally, the three types of content-analysis complement one another if the techniques of triangulation are used. Triangulation in content-analysis increases the validity and reliability of the outcomes (Oleinik 2010).

Content analysis is used in this case to address the principal research question as to how the Author’s text is read by the Reader. Namely, content analysis helps accomplish three tasks. First, with content analysis, the depth of reading is measured in a quantitative manner. How many ideas does the Reader identify in the text? Second, it makes a quantitative comparison of the Author’s and the Reader’s take on the same text possible. Are the Author’s ideas identified and interpreted by the Reader according to the former’s intentions? Third, what is the relationship between the constellation of words in a particular text and the ideas that the Readers manage to identify in it?

Case study of reading

Four social scientists were involved in the present study. Since their personal data was used, they will be referred to as A, B, C, and D for the sake of confidentiality. They worked independently in the past and have no co-authored publications (except the present one and the related one: Oleinik et al. 2014). Two social scientists have a mixed background in

economic sciences and sociology, one is a sociologist and one an applied social scientist (an expert in policy analysis). The first three participants are established scholars in the middle of their careers. Their works are cited actively in *eLibrary*, the largest Russian-language database of scholarly publications. The fourth participant is a young scholar.

The participants read samples of each other's texts published between 1999 and 2011 several times (4–5 times) and then performed a content analysis (see Hartley and Cabanac 2015 for a discussion of relevant research designs). The sample of A's works includes 20 articles, book chapters and book reviews, that of B's works—17 articles, book chapters and book reviews and that of C's works—20 articles and book reviews. In other words, A, B and C performed the roles of both Author and Reader. D played the role of Reader only, which allowed increasing the variability of the cases. Maximum variation cases help obtain information about the significance of various circumstances, namely, one's position in the chain of scientific communication (Fig. 1), for reading outcome (Flyvbjerg 2006, p. 230).

There were three stages in the content analysis of 57 texts. The design of each stage corresponds to a particular reading context. At Stage I, the participants read the texts and developed their codebooks (lists of qualitative codes) independently from one another. To assess the reliability of their qualitative coding, each reader created a dictionary based on substitution whose structure matched that of his/her codebook. After running three types of content analysis (qualitative, quantitative and with the help of a dictionary based on substitution), the distances between the texts in the three cases (measured by Cosine coefficients) were cross-correlated using an original method of triangulation (Oleinik 2010). A moderately-strong or strong association (measured by Pearson's r coefficients) was considered to be an indication of the reliability and validity of the Reader's qualitative coding.

At Stage II, the participants created a common codebook after several group discussions. At Stage I interactions between the Author and the Reader were mediated by the text only (without consulting each other on substantial issues of coding). In terms of game theory, the communication had features of a non-cooperative game. At Stage II, as negotiation and enforcement procedures became allowed, the context of reading grew closer to a cooperative game. Decisions had a consensual character: the Author and the Readers had the opportunity to propose new entries to the common codebook and to comment on the others' proposals. The common codebook contains 37 codes (15 entries for A's texts, 9 entries for B's texts and 13 entries for C's texts). Codes for A's texts were applied to A's texts only, those for B's texts to B's texts only, etc. Nevertheless, the participants independently recoded the texts using it.

The qualitative coding continued until an acceptable level of correlation was achieved between the three types of content-analysis—for every participant in particular and for the four participants in general. This time, the participants used a dictionary based on substitution created by common efforts.

No changes in the common codebook and the dictionary based on substitution were made at Stage III. However, all the codes were applied to all the texts. For instance, A's codes were used to content-analyze not only texts by A, but also those by B and C. This change was intended to explore whether the codes deriving from one Author's ideas provide useful insights for interpreting other Authors' ideas. For instance, the code 'Neoclassical theory' (in economics) is included in A's codes. However, the other scholars, particularly C, also discuss the theoretical assumptions and limits of the neoclassical economic theory, which paves the way for applying this code to C's texts. The above described reliability checks were also applied at Stage III.

Particular attention was devoted to studying the eventual association between the outcomes of the qualitative coding and of the quantitative content analysis (word co-occurrence). On the one hand, a particular combination of words found in a text determines the range of the possible: which ideas can be conveyed with their help. As stated in the “[Communicating at a distance in science](#)” section, this assumption is derived from the objectivist epistemological position. Consequently, by studying word co-occurrence one identifies “the range of things that speakers are capable of doing in (and by) the use of words and sentences” (Skinner 2002, p. 3). Word co-occurrence can be used as a proxy for the Author’s ideas: arguably, they determine the choice of words. A particular word can be used in a limited range of contexts. The word “to adjourn” has little currency outside the context of discussions of legal matters or that of talks about socialization in small groups, for example.

On the other hand, qualitative codes are generated and attributed to relevant fragments of a text in the process of reading. For instance, the common codebook at Stage III contains code “Paradigm” (key assumptions shared by most scholars working in a particular field; mentions of objectivist, subjectivist, evolutionary, systemic and anthropocentric paradigms broadly defined). This code was attributed to 33 fragments of A’s works, including the following excerpt: ‘(a particular theory) derives from the so called objectivist paradigm in sociology (E. Durkheim, T. Parsons and others)’. The coded fragments reflect how the Reader interprets the Author’s ideas and can be used as a proxy for the Reader’s take on the text. The subjectivist epistemological position is based on this assumption. By comparing the qualitative coding and word co-occurrence, one compares the Author’s and the Reader’s take on the same text and confronts its objectivist comprehension and constructivist interpretation at the same time.

The research design outlined above has several limitations. First, the four participants represent non-paradigmatic sciences. The level of agreement between social scientists is lower than that between natural scientists. Thus, a context cannot be confidently assumed to be shared by readers even if they read the same text (Hyland 2004, p. 32). Second, the level of agreement in the Russian social sciences tends to be particularly low. Scientific transactions in this case have a highly personalized and localized character. Personalized interactions prevail over depersonalized ones (Oleinik 2012). Third, and this is related to the second point, a number of the texts included in the sample were in the format of a scholarly essay. Compared with a standard scholarly article, essays might convey more than one idea and have a looser structure. Last, but not least, the small size of our sample—it only includes four cases—calls for moderation with respect to the interpretation of reported patterns.

The battle between author and reader

The title of this section should not be taken at face value. It refers to a game theory term. Namely, the title has its origins in the Battle of the Sexes. The Battle of the Sexes is a particular type of game, a nonzero-sum game characterized by a “mixture of conflict and mutual dependence” (Schelling 1960, p. 87). Their participants have divergent interests. When acting in their interests, nevertheless, the participants must adjust to one another (the Author to the Reader and vice versa). They gain together, but to an unequal extent, and they lose together, also to an unequal extent.

The subsections below refer to the consecutive stages in the content analysis.

Table 1 Reliability and validity of the reading at Stage I: Pearson correlation coefficients between the qualitative coding and word co-occurrence, 12 centroids (one centroid for each subsample of the texts and for each coder)

Reader (coder)	Author		
	A (N = 20)	B (N = 17)	C (N = 20)
A	.734**	.726**	.502*
B	.778**	.559*	.911**
C	.654**	.628**	.772**
D	.727**	.341	.644**

The highest coefficients of correlations for each Author (i.e. per column) are given in bold

* Refers to correlations significant at .05 level (2-tailed); ** refers to correlations significant at 0.01 level (2-tailed). The correlations are calculated between Cosine coefficients as measures of the distances between particular texts and the centroid. The texts are treated here as vectors of qualitative codes or words (e.g. a particular text contains n codes i , m codes j etc.; the same text contains l words a , k words b and so forth)

The author as an inattentive reader

The available evidence suggests that the Author reads his/her own texts less attentively than the Reader. It would be more accurate to say that the Author rereads them less attentively because he/she wrote and reread them several times. Judging by the strength of associations between the qualitative coding and word co-occurrence, none of the three Authors (A, B and C) managed to outperform the Readers. The Readers' qualitative coding was consistently closer to word-co-occurrence than the Author's (Table 1).

The fact that all of the Readers managed to achieve a satisfactory level of correlation between their qualitative coding and word co-occurrence also deserves attention. This suggests that the meanings discovered by the Readers lay within the space of possible readings determined by a particular constellation of words.

A comparison of the structure of the codebook developed by the Author for content analyzing his/her texts with the common codebook shows that the Author's codebook did not represent the best fit in a single case. The degree of structural homology between the participants' individual codebooks and the common codebook was measured with the help of the standard least squares fitting technique adapted to the circumstances of the present study. In all three cases (texts by A, B and C) the highest degree of structural homology was observed between D's individual codebook and the common codebook (Table 2).² The

² Table 2 contains the squared distances between the Reader's individual codebook and the common codebook. The method of least squares (the linear least squares fitting technique) is a standard approach to the approximate solution of over-determined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation (the OLS—Ordinary Least Squares criterion; Warner 2008, p. 52). In this particular case the common codebook was represented in the form of a 15×15 matrix for A's texts (because it contained 15 codes for content analyzing them), a 9×9 matrix for B's texts and a 13×13 matrix for analyzing C's texts. These "ideal" matrixes were compared with the matrixes obtained as a result of transforming the participants' individual codebooks. For instance, A's matrix referring to A's texts had 9 rows, which corresponds to the number of codes in A's individual codebook, and 15 columns, which corresponds to the number of codes in the common codebook. B's matrix referring to C's texts had 55 rows and 13 columns and so forth. Each cell contained either 1 (if a code from the Reader's codebook matches that from the common codebook) or 0. In the case of two perfectly identical matrixes, the sum of squared distances is equal to 0. In all other cases (there is a mismatch between the number rows or there is more than one code corresponding to a code in the common codebook), it exceeds 0. The formula for

Table 2 Degree of similarity between the ‘reading lenses’: Squared distances between the participant’s individual codebooks at Stage I and the common codebook developed at Stage II

Reader (coder)	Author			Total
	A (<i>N</i> = 20)	B (<i>N</i> = 17)	C (<i>N</i> = 20)	
A (<i>N</i> = 43*)	.537	.228	.163	.9275
B (<i>N</i> = 147)	.180	.2495	.230	.659
C (<i>N</i> = 70)	.102	.142	.60	.304
D (<i>N</i> = 44)	<i>.070</i>	<i>.102</i>	<i>.60</i>	.232
Mean	.222	.180	.128	.531

The least squared distance in each column (i.e. the highest homology with the common codebook) is given in italics. In the case of C’s works, C and D achieved a similar result

* Refers to the number of codes in the Reader’s individual codebook at Stage I

larger the reported value, the more dissimilar are a particular individual codebook and the common codebook and vice versa.

It should be remembered that D specialized on reading only in the present circumstances. The “perfect” Reader (because D’s take on the texts was not distorted by his/her authorship of some of them) produced the coding scheme that represents a compromise between the approaches of the other Readers. This does not mean, however, that D received the Author’s message with no distortions. At Stage I, D had relatively low scores (see Table 1), which suggests that D’s original reading did not necessarily catch the Authors’ ideas. In other words, D can be taught as a more representative Reader than A, B and C. D’s take on the texts arguably were positioned closer to that of the generalized Reader (the Reader who has neither familiarity with the Author’s texts nor personal connections with the Author).

To summarize findings presented in this subsection, the Author’s take on the text diverges from that of the generalized Reader, if they read it independently. Not only does the Reader’s interpretation depart from the Author’s line of reasoning as a result of the former’s particular interests, background knowledge, cognitive capacities, etc., but the Author is not necessarily able to identify all the ideas and concepts that the text may eventually suggest. In more literary terms, the text exists separately from the Author and his/her take on it. The Author may read it differently as time passes.³

Footnote 2 continued

calculating the sum of squared distances contains the following components: $\sum (M - \bar{M})^2 = \sum_{i=1}^n (x_i - \bar{x}_i)^2 + \sum_{j=1}^m (x_j - \bar{x}_j)^2$, where *M* refers to a particular Reader’s matrix, \bar{M} —to the common matrix (with the common codebook at its origin), x_i —to the column *i* mean in *M*, \bar{x}_i —to the column *i* mean in an “ideal” matrix perfectly matching \bar{M} , x_j —to the row *j* mean in *M*, \bar{x}_j —to the row *j* mean in \bar{M} , *n*—to the number of columns in *M*, *m*—to the number of rows in *M*. \bar{x}_i and \bar{x}_j depends on the number of codes in the common codebook. For instance, if the common codebook contains 5 codes, an “ideal” matrix perfectly matching it has 5 columns and 5 rows: $\bar{x}_i = \bar{x}_j = \frac{1}{5} = 0.2$ (because each column and row contains one non-empty cell).

³ The same arguably applies to the Reader, but the present research design does not allow capturing changes in the Reader’s take on a text over time.

Comeback of the author

After producing the common codebook, the participants reread the texts at Stage II, applying the same set of codes to the corresponding subsample: the A codes—to A's texts, the B codes—to B's texts and the C codes—to C's texts. The Stage II design served to compare the readings of the same text, on the one hand, by the Author and, on the other hand, by the Reader when their task consisted in identifying the same ideas and concepts. Does the Author interpret the text more closely to the range of meanings determined by a particular constellation of words than the Reader?

The Stage III design had a different rationale. It was intended to test the assumption that a set of codes adapted for the interpretation of one Author's works provide little support for reading texts by the other Authors. Can B's and C's texts be read with the help of the A codes? The participants reread all the texts one more time applying the codes to all of them. Namely, the A codes—to A's, B's and C's texts and so forth. The Stage III design also serves to test the limits of the objectivist comprehension. If a set of codes that derives from a particular text can also guide the reading of other texts, then the objectivists' belief in the embeddedness of meanings in the text would be undermined.

The outcomes of Stages II and III show that the Author's qualitative coding tended to be more closely associated with word co-occurrence than that of the Readers (Table 3). This finding suggests that the Author outperformed the Reader when identifying meanings encoded in a particular constellation of words. There was one exception only: D's readings of C's texts at Stage II (D's qualitative codes represented a better match to word co-occurrence in C's texts). Nevertheless, the Readers managed to achieve a satisfactory level of correlation between their qualitative coding and word co-occurrence. Once again, the meanings attributed by them lay within the space of the possible readings determined by a particular constellation of words.

The Author decides the parameters of the constellation of words when writing the text. He/she also appears to be better prepared than the Reader to interpret it using a given template. The last condition—the application of a given template for reading—has to be particularly emphasized as it helps explain an apparent contradiction with the finding reported in the previous subsection (the Author's failure to identify *all* the meanings that his/her text contains). The Author may well omit *some* meanings suggested by his/her texts. But he/she is good at identifying a *restricted range* of meanings.

Speaking more specifically about the outcomes of Stage III in the qualitative coding, they confirm that a coding scheme developed for interpreting one Author's works had a limited applicability beyond his/her texts. First, the level of inter-coder agreement decreased when the codes that derive from one Author's texts were applied by the same Readers to texts written by the other Authors. The Krippendorff's Alpha values for each pair of the coders varied from .496 to .575 at Stage II and from .399 to .412 at Stage III.⁴ The Readers disagreed more often when looking at a text through the "improper lens".

⁴ When assessing these values of Krippendorff's α , one has to bear in mind that the suggested cut-off points, .667 and .800 (Krippendorff 2004, p. 241), do not take into account, on one hand, an inverse relationship between the values of α and the number of coding categories, 37 in the present case (Muñoz-Leiva et al. 2006; Oleinik et al. 2014) and, on the other hand, the fact that coded units are not given or natural here. The Readers applied codes to particular segments of the text, as opposed to the text as a whole (values of Krippendorff's α are higher in the latter case). No known computer program allows calculating the coefficients of agreement separately for coding and for unitizing where unitizing refers to partitioning of a given continuum into sections (Krippendorff 2004, pp. 251–256; see also Oleinik 2010, p. 872).

Table 3 Reliability and validity of the reading at Stages II and III: Pearson correlation coefficients between the qualitative coding and word co-occurrence, 3 centroids (one for each subsample) at Stage II and one centroid at Stage III

Reader (coder)	Author					
	Stage II			Stage III		
	A (N = 20)	B (N = 17)	C (N = 20)	A (N = 20)	B (N = 17)	C (N = 20)
A	.600**	.445	.701**	.438	.117	.811**
B	.562**	.484*	.625**	.429	.498*	.824**
C	.542*	.458	.693**	.361	.437	.865**
D	.410	.484*	.917**	.360	−.135	.734**

The highest correlation coefficients in each column are given in italics. Values that do not fit the general pattern are given in bold (e.g., C was expected to be the best reader of C’s works using the common codebook)

Second, an inspection of the distribution of fragments coded by the Author and by the Coder suggested that the overwhelming majority of the cases were arrayed on a diagonal: 83.1% of the fragments coded with the help of the A codebook referred to A’s texts, 88.6% of those coded with the help of the B codebook—to B’s texts and 84.7% of those coded with the help of the C codebook—to C’s texts. The existence of this pattern was further confirmed by high values of the Chi square statistics and Lambda: $\chi^2 = 7393.8$, $df = 4$, significant at $p < .001$; $\lambda = .73$ (dependent variable: Codebook).

Can one deduce that ideas and concepts tend to have an Author-specific character without contradicting the conclusion of the previous subsection about the multiplicity of the readings of a text? This is possible since at Stage III very specific types of templates for reading were applied to the Author’s texts. Instead of resulting from a Reader’s interests, background knowledge and cognitive capacities, they referred to the particularities of the reading of the other Author’s texts. In other words, a limited usefulness of a particular set of ideas and concepts that derives from the other Author’s works for the interpretation of the Author’s text does not preclude that the Readers are able to find a potentially unlimited number of ideas and concepts in it.

To conclude, the discussion above shows that, in the case of the individual reading, the Author’s take diverges from the Reader’s take. The Author does not have an exclusive control over his/her creation. However, the Author does a better job than the Reader in the case of interpreting his/her texts along particular lines. The closer these lines are to a constellation of words at the origin of the Author’s text, the clearer the Author’s advantage over the Reader is.

Limited capacity to read

A Reader’s limited cognitive capacities, coupled with the proliferation of scientific publications, prevent him/her from identifying all possibly interesting ideas in a text. The depth of reading in science is rarely, if ever, empirically measured (except in very specific cases of standardized tests of reading comprehension). The qualitative coding provides an opportunity for measuring the depth of the reading in a quantitative manner. Namely, by comparing the number of codes and coded segments at Stages I and II, one is able to assess the depth of the reading in various circumstances (using the Author’s individual codebook,

Table 4 Depth of the reading: number of codes and coded segments at Stage I and II and the *eLibrary* data in respect of the selected texts

Reader (coder)	Author					
	A (<i>N</i> = 20 [8*])		B (<i>N</i> = 17 [13])		C (<i>N</i> = 20 [11])	
	Codes	Segments	Codes	Segments	Codes	Segments
<i>Stage I</i>						
A	8	1863	17	1082	18	1509
B	57	1123	35	540	55	836
C	30	1089	16	712	24	712
D	19	593	11	681	14	688
Mean (readers)	35	935	14.7	825	29	1011
<i>Stage II</i>						
A	15	433	9	244	13	283
B	15	616	9	301	13	376
C	15	580	9	384	13	333
D	15	527	9	399	13	366
Mean (readers)	15	574.3	9	342.3	13	341.7
<i>eLibrary</i>	6	10 [16**]	6	22 [29]	5	14 [34]

Values for the Author reading his/her own texts are given in bold

* Refers to the number of this Author's texts included in *eLibrary*; ** refers to the total number of citations of this Author's texts in *eLibrary*, including non-comprehensive ones

using the Reader's individual codebook, using the common codebook). In addition, the primary data were complemented by the secondary data about A's, B's and C's publications included in *eLibrary*, namely, the number of comprehensive references to them. The comprehensive reference (excluding self-references) is defined as a detailed discussion of the Author's ideas, methods or data that goes beyond a simple mention of his/her work. The comprehensive references were content analyzed using the common codebook developed at Stage II.

The number of codes in the Author's and Reader's individual codebooks, as well as the number of fragments coded at Stage I, consistently exceeded the corresponding figures at Stage II (Table 4). Regardless of the coders' individual styles (some coded far more fragments than the others at Stage I), they identified fewer significant fragments when using the common codebook than when using the individual codebooks. The coders' own texts were no exception in this respect.

Judging by the number of the comprehensive references to A's, B's and C's works, the Russian Readers in general (citing Authors) interpreted these texts in an even narrower manner. For instance, the Russian readers comprehensively discussed C's ideas and concepts on 14 occasions referring to 5 codes from the common codebook. As a comparison, C's individual codebook for analyzing C's texts contained 24 codes. C identified 712 fragments corresponding to these 24 codes at Stage I. The common codebook for analyzing C's texts includes 9 codes. C identified 333 fragments corresponding to these 9 codes at Stage II. This means that the Russian scholars found 5 codes from the common codebook (38.5%) relevant to their own interests and background knowledge. They applied these codes in the same manner as the Author did on 14 occasions (4.2%).

Similar figures for the other participants are: 40% (6 codes out of 15) and 2.3% (10–433 fragments) for A, 66.7% (6 codes out of 9) and 7.3% (22–301 fragments) for B. On the one hand, these figures represent a very rough (because the other ideas developed in the citing Authors' texts are not taken into account) approximation for the degree of agreement between the attentive Readers in general and the four attentive Readers in particular (A, B, C and D) with respect to the significance of the particular ideas and concepts. The attentive Readers in general used from one- to two-thirds of the entries in the common codebook only. On the other hand, they are indicative of the depth of the reading done by the attentive Readers in general relative to the depth of the Author's reading. In general, the attentive Readers identified less than 10% of the fragments that refer to the significant ideas and concepts.

By any account, even the attentive Reader uses only a few ideas and concepts that the text may actually contain. It takes time and cognitive resources for a Reader to become an attentive Reader. A, B, C and D read the 57 texts at least 4 times (a preliminary reading plus three readings at Stages I, II and III) spending from 15 min to 4 h on reading and coding particular texts, depending on their length, readability and the stage. A's evaluation of the total time spent on the readings and coding is 120 h plus, B's evaluation—250 h plus, C spent about 100 h, and D spent about 250 h.

The inattentive Reader makes still less use of the text, settling for a perfunctory reading of it. The majority of scholars do not reread texts, however. Even a second reading is rare. According to one study, only 20% of the faculty members of an Australian university actually reread texts (Wilson and Tenopir 2008, p. 1403). This means that, even when the Author is read, which is itself a rare outcome of textually mediated communication at the time of proliferation of scholarly publications, this does not guarantee that the Reader's will interpret the Author's ideas and concepts according to the Author's intentions.

Does the reading ease help?

Can a Reader's limited capacity to read be alleviated by the Author's production of more readable texts? The current research design does not allow directly answering this question. Nevertheless, some findings may be relevant for subsequent inquiries in this matter.

The Flesch Reading Ease measure is an imperfect (Hartley 2016) yet still widely used proxy for the readability of text. A very strong association was found between the mean readability scores of the texts ($M_{\text{Flesch}} = -69.36$, $\sigma = 12.88$ for A, $M_{\text{Flesch}} = -62.96$, $\sigma = 12.49$ for B and $M_{\text{Flesch}} = -53.48$, $\sigma = 8.06$ for C⁵) and the mean squared distances reported in the last row of Table 2. The value of Pearson's correlation coefficient is $r = -.999$, $p = .016$ (1-tailed), $N = 3$. Despite the smallness of the sample, this finding tentatively suggests that the more readable the text is, the easier for its Readers to reach an agreement. It should be borne in mind, however, that an agreed reading does not necessarily correspond to the Author's intentions.

The study of partial correlations between the qualitative coding and word-occurrence controlling for the Flesch scores provides some additional insights. In most cases (7 out of 12), the correlation decreased, but did not drop significantly, $r_{\text{cq}} > r_{\text{cq.F}}$. This means that

⁵ This means that C's texts tend to be relatively more readable and A's texts relatively less readable. The use of the formula for calculating the Flesch score that was not specifically adapted to Russian language does not prevent one from assessing the *relative* readability of texts. For instance, the mean readability score for a non-random sample of articles in Russian available at three major news websites was -32.03 ; the mean readability of Leo Tolstoy's tales for children was $+23.51$.

the readability of texts partially *mediates* the association between the qualitative coding and word co-occurrence (Warner 2008, pp. 407–409). The more readable texts written by A and B were, the more chances exist that they were interpreted according to the Author’s intentions. In the cases when D reads B’s texts and C reads C’s texts, the readability turned out to be irrelevant: $r_{cq} \approx r_{cq,F}$. Once again, one must be very cautious when attempting to make any generalizations.

Finally, readability did not have an impact on the number of references in *eLibrary*: $r = .058$, $N = 32$. At the same time, the number of figures and tables—they increase the clarity with which a text explains what it says (Stremersch et al. 2007, p. 174)—turned out to be associated with the number of references in *eLibrary*: $r = .499$ at $p = .004$, $N = 32$ ($r = .849$ at $p < .001$, $N = 32$ in the case of the comprehensive references).

Conclusion

Our study explores some particularities of text-based peer-to-peer communication in science that refers to an under-researched area in the sociology of science. To address the research questions formulated in the introduction, we use an original methodological approach that combines qualitative and quantitative data on academic reading.

The first question refers to how the Author’s text is read by the Reader. When the Reader and the Author both read the Author’s text, the Author appears to be better at searching for and identifying his/her own ideas. Compared with the Reader, the Author interprets his/her texts with the help of a *given* set of ideas and concepts in a more valid and consistent manner. This conclusion, which appears trivial at first, does nevertheless serve to indicate that the Author should not hope to be properly understood even if he/she specifically highlights key ideas. We approximate this situation by developing a common codebook at Stage II. The common codebook was expected to direct the Readers’ attention to the same ideas. Our findings warn against making the ambitious conclusion that, if the Author carefully discusses his/her ideas at length, the Reader will necessarily grasp them. This may or may not happen. Even in the case of our quasi-laboratory experiment with controlled communication between the Author and prepared Readers we reported some loss of the information that the latter intended to convey to the former.

The second question is to what extent does the reading of the Author’s text allow the Reader to find new ideas that do not necessarily go along with the Author’s train of thought. The Authors’ qualitative coding at Stage I (the participants read the texts independently) was not the closest to the constellations of words, as shown in Table 1. The Authors’ individual codebooks (“reading lenses”) did not represent the closest match to the common codebook at Stage II either, as shown in Table 2. The *given* set of ideas and concepts does not necessarily derive solely from the Author’s intentions.

Both the Author and the Reader read the text in a either deep or perfunctory manner. Even the Author might need—as time passes—to invest significant time and cognitive resources into interpreting his/her own texts again. Consequently, the deep reading cannot be “easy” as it requires the Reader’s continuous attention and willingness not to take any particular interpretation for granted. The rule of “the smallest processing effort” formulated in relevance theory (White 2011, p. 3347) might facilitate a perfunctory reading. In the case of a deep reading, nevertheless, the more time and attention invested, the more value-added (interpretations) it produces.

Deep reading makes two-way text-based communication, namely a dialogue between the Author and the Reader, possible. Our study suggests that deep reading (as opposed to perfunctory reading) leads to the discovery of new ideas in addition to the ideas that the Author initially intended to convey. The attentive Reader (including the Author as a Reader) receives more information than the Author intended. The Reader finds in the Author to be either an ally (if the reading has a cooperative character) or a foe (if the reading has a critical character) when developing the Reader's own ideas.

While, in the previous case, we reported losses of information in text-based communication, the information was actually enriched in the case at hand. The information is enriched even from the Author's point of view because the two-way communication helps the Author adapt a new perspective for looking at his/her own text. The two-way, text-based communication constitutes an important source of new ideas in science. This optimistic finding contrasts with the rather pessimistic conclusion about the loss of information in the process of reading.

Our study also provides some arguments that are relevant for a more general discussion of the epistemological foundations of science communication. Namely, the objectivist comprehension and the constructivist interpretation do not necessarily represent absolute antipodes. They are both complementary and mutually exclusive. The principle of objectivist comprehension determines the space of the possible readings of a text as a constellation of specific words. However, the choice of a particular reading lens *within* this space has a subjective component and may be better studied in terms of constructivist comprehension.

Further explorations of textually mediated communication could be made along the following axes. First, in order to gain further insights, it would be necessary to replicate the study in the context of a paradigmatic science characterized by a stronger consensus between the scholars. Second, and this is related to the first, a replication of the study using highly structured standard articles only would serve to determine if this format of scientific communication restricts the range of the possible interpretations of the Author's ideas. Third, the impact of the readability of scholarly texts on their comprehension and interpretation should be analysed more systematically.

Acknowledgements The authors are grateful for the Scientometrics anonymous reviewers' deep and constructive reading of the text. An earlier, Russian, version of the text appeared in *SOCIS* (Sotsiologicheskie Issledovaniia). The remaining errors and inaccuracies are their own.

References

- Abbott, R. (2013). Crossing thresholds in academic reading. *Innovations in Education and Teaching International*, 50(2), 191–201.
- Auerbach, B. (2006). «Publish and Perish»: La définition légitime des sciences sociales au prisme du débat sur la crise de l'édition SHS. *Actes de la recherche en sciences sociales*, 164, 75–92.
- Bakhtin, M. (1979). *Problemy poetiki Dostoevskogo* [Problems of Dostoevsky's Poetics] (4th ed.). Moscow: Sovetskaya Rossiya.
- Bourdieu, P. (1984). *Homo academicus* (Paris ed.). Paris: de Minuit.
- Brossard, D., & Shanahan, J. (2006). Do they know what they read? Building a scientific literacy measurement instrument based on science media coverage. *Science Communication*, 28(1), 47–63.
- Budd, J. M. (2001). Misreading science in the twentieth century. *Science Communication*, 22(3), 300–315.
- Fischhoff, B. (2013). The sciences of science communication. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3), 14033–14039.

- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245.
- Goldgar, A. (1995). *Impolite learning: Conduct and community in the Republic of Letters, 1680–1750*. New Haven: Yale University Press.
- Gryaznova, Yu., & Ratz, M. (2008). Kak nam chitat'?: [How shall we read?]. *Voprosy psikholingvistiki*, 8, 142–146.
- Hartley, J. (2006). Reading and writing book reviews across disciplines. *Journal of the American Society for Information Science and Technology*, 57(9), 1194–1207.
- Hartley, J. (2016). Is time up for the Flesch measure of reading ease? *Scientometrics*, 107(3), 1523–1526.
- Hartley, J., & Cabanac, G. (2015). An academic odyssey: Writing over time. *Scientometrics*, 103(3), 1073–1082.
- Hartley, J., Sotto, E., & Fox, C. (2004). Clarity across the disciplines: An analysis of texts in the sciences, social sciences, and arts and humanities. *Science Communication*, 26(2), 188–210.
- Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41(3), 497–518.
- Henige, D. P. (2006). Discouraging verification: Citation practices across the disciplines. *Journal of Scholarly Publishing*, 37(2), 99–118.
- Hirschauer, S. (2010). Editorial Judgments: A praxeology of “Voting” in peer review. *Social Studies of Science*, 40(1), 71–103.
- Hultberg, J. (1997). The two cultures revised. *Science Communication*, 18(3), 194–215.
- Hyland, K. (2004 [2000]). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: The University of Michigan Press.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.
- Latour, B. (1987). *Science in Action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Littmann, M. (2005). Courses in science writing as literature. *Public Understanding of Science*, 14(1), 103–112.
- Muñoz-Leiva, F., Montoro-Ríos, F. J., & Luque-Martínez, T. (2006). Assessment of interjudge reliability in the open-ended questions coding process. *Quality and Quantity*, 40(4), 519–537.
- Norris, S. P., & Philips, L. M. (1994). The relevance of a reader's knowledge within a perspectival view of reading. *Journal of Reading Behavior*, 26(4), 391–412.
- Oleinik, A. (2010). Mixing quantitative and qualitative content analysis: Triangulation at work. *Quality and Quantity*, 45(4), 859–873.
- Oleinik, A. (2012). Publication patterns in Russia and the West compared. *Scientometrics*, 93(2), 533–551.
- Oleinik, A., Popova, I., Kirdina, S., & Shatalova, T. (2014). On the choice of measures of reliability and validity in the content-analysis of texts. *Quality and Quantity*, 48(5), 2703–2718.
- Sapir, E. (1961). *Culture, language and personality*. Berkeley: University of California Press.
- Schelling, T. S. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review*, 68(2), 2–16.
- Skinner, Q. (2002). *Visions of politics: Regarding method* (Vol. 1). Cambridge: Cambridge University Press.
- Spektor-Levy, O., Eylon, B.-S., & Schrez, Z. (2009). Teaching scientific communication skills in science studies: Does it make a difference? *International Journal of Science and Mathematics Education*, 7(5), 875–903.
- Stanovich, K. E. (2003). Understanding the styles of science in the study of reading. *Scientific Studies of Reading*, 7(2), 105–126.
- Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3), 171–193.
- Volentine, R., & Tenopir, C. (2013). Value of academic reading and value of the library in academics' own words. *Aslib Proceedings: New Information Perspectives*, 65(4), 425–440.
- Warner, R. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.
- Weber, M. (1968). *Economy and society: An outline of interpretative sociology*. New York: Bedminster Press.
- Weigle, S. C., Yang, W. W., & Montee, M. (2013). Exploring reading processes in an academic reading test using short-answer questions. *Language Assessment Quarterly*, 10(1), 28–48.
- White, H. D. (2011). Relevance theory and citations. *Journal of Pragmatics*, 43(14), 3345–3361.
- Whorf, B. L. (1956). *Language, thought and reality: Selected writings*. Cambridge, MA: MIT Press.

- Wilson, C. S., & Tenopir, C. (2008). Local citation analysis, publishing and reading patterns: Using multiple methods to evaluate faculty use of an academic library's research collection. *Journal of the American Society for Information Science and Technology*, 59(9), 1393–1408.
- Yore, L. D., Hand, B. M., & Florence, M. K. (2004). Scientists' views of science, models of writing, and science writing practices. *Journal of research in science and teaching*, 41(4), 338–369.
- Yore, L. D., Hand, B. M., & Prain, V. (2002). Scientists as writers. *Science Education*, 86(5), 672–692.