

University of Groningen

Literary quality in the eye of the Dutch reader

Koolen, Corina; van Dalen-Oskam, Karina; Cranenburgh, van, Andreas; Nagelhout, Erica

Published in:
Poetics

DOI:
[10.1016/j.poetic.2020.101439](https://doi.org/10.1016/j.poetic.2020.101439)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Koolen, C., van Dalen-Oskam, K., Cranenburgh, van, A., & Nagelhout, E. (2020). Literary quality in the eye of the Dutch reader: The National Reader Survey. *Poetics*, 79, [101439].
<https://doi.org/10.1016/j.poetic.2020.101439>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Literary quality in the eye of the Dutch reader: The National Reader Survey



Corina Koolen^{a,*}, Karina van Dalen-Oskam^{a,b}, Andreas van Cranenburgh^c,
Erica Nagelhout^d

^a Huygens Institute for the History of the Netherlands, Amsterdam, the Netherlands

^b University of Amsterdam, Amsterdam, the Netherlands

^c University of Groningen, Groningen, the Netherlands

^d GfK, Amstelveen, the Netherlands

ARTICLE INFO

Keywords:

Readers
Literary quality
Popular fiction
Bestsellers
Survey

ABSTRACT

What makes some novels *literary*? There is little agreement within literary studies on this question. The two main approaches focus either on text-intrinsic factors (e.g., aesthetic, stylistic), or text-extrinsic social factors (e.g., author prestige, critics). Until now, there has not been a comprehensive study taking both text-intrinsic and social factors into account. The project The Riddle of Literary Quality examines both factors by connecting literary texts to the appreciation of those texts: can we identify textual characteristics that are connected to readers' literary appraisal of texts? In this paper, we describe the development of The National Reader Survey and present some results. The National Reader Survey is a large online survey of Dutch readers with about 14,000 respondents; its purpose is to collect readers' literary appraisal of texts. We asked readers to rate both read and unread novels on a scale of 1–7 for their literary and overall quality. The agreement amongst respondents on which recent Dutch language novels are of high literary quality and which are not was greater than expected. Motivations respondents gave for their ratings show that the notion of literary quality is a familiar one and respondents most commonly relate it to two elements: the first is the text itself—style, structure, plot and layers; and the second is genre—if a novel is considered a 'genre' novel (e.g., suspense, romantic, fantasy), its chances of obtaining a high rating on literary quality are small. These results indicate how entwined social and text-intrinsic factors are. We also touch upon project results which make use of the survey data.

1. Introduction

Debates about literary quality involve not just what the term precisely constitutes, but also whether it is even a legitimate concept to begin with. Sociologists have claimed the concept to be basically void of meaning (e.g., Bourdieu, 1984), whereas feminists argued against the exclusionary function it performs (e.g., Ellmann, 1968). Notwithstanding such theoretical considerations, the practice of applying notions of literary quality is still very common in the appraisal of fictional works, but its characteristics are not well understood. Empirical research on the topic is generally focused on the workings of specific elements of literary quality, such as

* Corresponding author at: Huygens Institute for the History of the Netherlands, Oudezijds Achterburgwal 185, 1012 CJ Amsterdam, the Netherlands.

E-mail address: corinakoolen@gmail.com (C. Koolen).

<https://doi.org/10.1016/j.poetic.2020.101439>

Received 18 October 2018; Received in revised form 22 January 2020; Accepted 26 January 2020

Available online 15 February 2020

0304-422X/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

foregrounding, but not on the notion itself. The digital humanities project The Riddle of Literary Quality aims to address this by examining two aspects of literary quality: the existing perceptions of literary quality, as well as the actual texts of novels and how these two aspects are connected. We do this by applying a bottom-up empirical approach. Rather than pre-select aspects of literary quality (as defined in literary theory) and examine these, or pre-select texts regarded as of high literary quality (by literary critique or literary theory, for instance), the project team decided to construct and conduct a survey asking readers from the general public for their opinions on a wide range of recent fictional novels.

The project's ultimate aim is to examine the relationship between the *text* of prose fiction novels and their perceived level of literary quality. In other words, we ask ourselves the question: to what extent can we retrace perceptions of literary quality back to textual characteristics through computational means? In order to research texts perceived to be of different literary levels, we needed empirical data on literary quality. To this end, the project devised a reader survey, *Het Nationale Lezersonderzoek* (The National Reader Survey), to obtain information on how literary quality is perceived among the general Dutch reading audience and specifically, how readers judge certain novels. This survey is the topic of this article.

Not only did we ask respondents to rate recently published novels on a scale of overall and literary quality, we also obtained information from respondents on the type of reader they are (for instance, do they read more for recreation or for aesthetic pleasure?). We also asked respondents to explain the motivation for one of their ratings in order to examine which criteria respondents use.

This article is structured as follows. First, we discuss the theoretical background of the survey (Section 2); next, we discuss the method (Section 3), including the pilot study undertaken beforehand to see how the survey would function and its results (Section 4). The results are diverse: we collected information on reader roles (Section 4.1), ratings of the general and literary quality of novels read, and also ratings of the general and literary quality on novels respondents had not read. We did this to assess the influence of social factors on the ratings (Section 4.2). We then discuss the written motivations on literary quality (Section 4.3). Finally, we briefly touch upon the results of two PhD theses that used the results of the survey (Section 5).

2. Theoretical framework

First, we define the central concept of this paper, literary quality, and then position our work in relation to existing studies.

2.1. Concept: literary quality

As a concept, literary quality is important to examine because judgments about it are still used to include or exclude works for serious discussion in the literary field, and ultimately, the literary canon (see, for instance, Van Peer, 2008). We define literary quality as “a consensus, influenced by the literary establishment, which consists of a combination of extra-textual and textual qualities” (Koolen, 2018). We consider the literary establishment to be a network of professionals who decide which literary works are published, reviewed and awarded prizes (cf. Dorleijn & van Rees, 2006: 15). An example of an extra-textual quality is genre. Even though one could argue that genre is reflected in text, readers are steered in their perception of genre before reading, based on the publisher's marketing of novels (for instance, take the book cover: title, font, use of images and colors). As we shall see, certain genres are automatically excluded from positive literary quality judgments. Textual features are, for instance, writing style, which includes sentence length and word usage. This definition implies that we believe there are aspects of literary text that contribute to it being perceived as having high literary quality. But it also means that this is not a static consensus nor a universal one, nor necessarily a socially fair one, either—think of the bias towards gendered, racial or social class-related preferences (Robinson, 1983). It is situated temporally and geographically and is strongly related to aesthetic considerations, albeit not exclusively so. We examine Dutch-language prose fiction, both originally written in Dutch as well as translated into Dutch, from a particular time frame: works published between 2007–2012, which were best selling or borrowed most often from libraries in the period 2009–2012. Therefore, we can only make claims about this specific language, this type of novel and this specific time frame, though the survey can be repeated to take a different measurement.

In the project, we apply a different approach to the analysis of literary quality than in literary-theoretical analysis or in literary critique. Rather than define a set of characteristics beforehand and then look for these characteristics in novels, we use the judgments of the general reading public on which novels they regard as having high literary quality, low literary quality, and everything in between.

Finally, from the perspective of feminism and postcolonial studies, the notion of literary quality and its role in canon formation has been critiqued. Robinson (1983) provides a thoughtful and detailed analysis of the problems with canon formation and questions of literary quality encountered in feminist attempts to open up the canon to female, black, and lesbian literature; a conundrum that has not been solved to date. Said (1993) shows the influence imperialism still has on culture. Both show that canon formation is never value-free and continues to replicate colonial structures, which results in the exclusion of works by less privileged people. Our research should not be taken to call such critical thought into question—this continued inequality is still a factor in the reception of novels as more or less prestigious.

2.2. Related work

What constitutes literary quality is a long-standing debate, both in literary theory and in other (non)academic disciplines. In this section, we will discuss some disciplines that examine the concept, as well as readers' appreciation of fictional works in general.

In non-empirical literary studies, the discussion swings between two main approaches. The first holds that the quality of a literary text is a result of features being foregrounded, for instance, metaphorical language, resulting in the defamiliarization of the reader (Sjklowski, 1917; Mukarovsky, 1940; Jakobson, 1960). The second approach claims that texts are valued according to sociological processes in the literary field, positioning certain authors at the literary forefront and others on the fringes (Bourdieu, 1984; De Nooy, 2002; Dorleijn & van Rees, 2006; Fokkema & Ibsch, 1992; Rees & Dorleijn, 1993; Verboord, 2003). The formalist and structuralist scholars from the first approach focus on the texts themselves and their prosodic, grammatical, and semantic characteristics. The literary sociologists who favor the second approach analyse the workings of literary institutions and the reputation of authors.

In the research program that Bourdieu inspired, intrinsic features of the artwork or text receive little attention. This is seen as a serious drawback in the “new sociology of art.” Scholars such as Hennion (2007) try to bring the objects themselves back into the sociological picture. They state that taste is a choice and is consciously or unconsciously developed during one’s life. Hennion uses wine tasting as an example and focuses on amateurs instead of professionals. Amateurs may follow different trajectories but end up acquiring the same tastes as professionals. Schwarz (2013) follows up on this, focusing on the differences in tasting techniques in the observers of art or readers of books, and hypothesizing that these differences may still relate to differences in sociological background of the observers and readers (and respectively wine tasters). Our point of departure in *The Riddle of Literary Quality* resembles the perspective taken by Hennion and Schwarz, paying attention to both amateur and professional. Theorists in cultural studies have suggested that there is a trend where the sharp boundary between high and low culture is fading (e.g., Peterson, 1992; Sontag, 1994). For the survey we present in this contribution, this informed our choice in using a corpus consisting of titles with a mostly high economic value, but also containing enough titles regarded as having sufficient literary prestige to make a useful comparison.

It is only since the 1960s that readers have been taken into account as a potentially important factor in determining literary quality (Barthes, 1968; Iser, 1970, 1974 Jauss, 1967). Jauss (1967) argues that novelty in literary texts caused by deviation, defamiliarization and estrangement affects literary values and evaluation (Schram, 1985, p. 7). Research in experimental aesthetics investigates these *foregrounding* effects (see, for instance, Da Costa Fialho, 2007; Fricke, 2008; Miall & Kuiken, 1994; Miall & Kuiken, 1999; Hakemulder, 2004; Van Peer, Hakemulder, & Zyngier, 2007), but also readers’ absorption into the text (Hakemulder & Kuijpers, 2017). The results also provide insight into the differences in the reading process of literary fiction versus non-fiction. While the research on the effects of foregrounding elements on the literary reading process has yielded indispensable empirical results, the experimental setup in which participants are closely monitored during their reading process is too time-consuming to employ on a large corpus of complete texts.

In one branch of experimental aesthetics based on the ideas of Fechner (1876), the subjects’ ideas on aesthetics are researched. Respondents are asked to recall as many adjectives related to aesthetic judgment as possible, for instance of literary novels. In Knoop, Wagner, Jacobsen, and Menninghaus (2016), this approach leads to the conclusion that ‘beautiful’ and ‘suspenseful’ are central to descriptions of aesthetic judgment of fictional literature. This type of research is related to the influence of the text, but such studies do not ask readers to reflect on specific works. More generally, there is more to literary quality than aesthetics. Even though there has been a focus on the aesthetic experience of literary language (see Van Peer, 2008), ethical, moral, affective and other motivations can be part of literary judgment as well.

Finally, in recent years, attention has turned to the online engagement of readers of fiction and its influence on reception; e.g., Chevalier and Mayzlin (2006), Verboord (2011), Vlieghe et al. (2016), Boot (2013), and Thelwall (2017); which all give examples of how book review platforms such as GoodReads influence literary production and consumption. Many online reviews are readily available on a wide range of genres, especially on bestselling fiction. However, we opted not to use existing online databases as we are specifically interested in literary quality across genres, and literary quality is typically not discussed in reviews of novels that are seen as not very literary.

3. Method

3.1. Motivation and aims

The premise behind *The Riddle of Literary Quality* project is to investigate the association between textual characteristics and the appraisal of the literary quality of novels (<http://literaryquality.huylens.knaw.nl>). In other words: can ideas about the literary quality of a novel—high, low, or somewhere in between—be connected to the actual text of the novel? However, it is then necessary to have a consistent measure for the literary quality of those novels; hence we made a number of decisions, which we will explain briefly here.

First: whose opinions should we use or collect and how? We decided against relying solely on the opinions of professional critics (or our own), because we are interested in the reception of novels by the general reading public, and because we needed a sufficient number of opinions in a similar format and on a sufficient number of novels to be able to make reliable generalizations from our corpus of novels. Thus, the first decision was to collect reader judgments.

We did not mine ratings from review websites because we wanted to be able to distinguish between literary quality and overall quality—for instance, a novel might be good, but not necessarily literary. ‘Star’ ratings commonly used in reviews do not make a distinction between the two. Moreover, we wanted to have more information on the readers’ literary preferences. Therefore, we decided to devise a survey and ask readers directly for ratings on a scale of literary quality and on a scale of overall quality. Note that we did not supply respondents with a definition of the concepts, precisely because we wanted to explore ideas in the current Dutch reading population without steering anyone towards socially desirable answers. To ensure these ratings were not based mainly on what readers remembered as ‘appropriate’ novels for high school reading lists, we decided to ask participants to rate a list of recently

published novels. By restricting this list to successful novels, we aim to realize the goal of obtaining a sufficient number of ratings.

Before conducting the actual survey, we undertook a panel survey. One of the reasons for this was to estimate the number of participants needed to obtain a sufficient number of ratings for each novel. This resulted in an aim of 18 100 participants to start the survey, knowing that not everyone would complete it (see Section 3.2 for further explanation). It was an ambitious aim which we hoped to reach by making sure the survey was publicly accessible for a longer period of time, and by spending part of our budget on public relations activities (see Section 3.3).

3.2. Pilot study

The data collection for the pilot study (in the form of a panel survey) and the final survey was performed by NoTies, which was responsible for the adaptation of its survey-specific software to our needs and for hosting the survey and exporting the results for further analysis. After several discussion sessions, we arrived at a pilot version of the survey. We had this version tested by a panel of 1130 Dutch citizens. The service we used for this test is the Invotes panel owned by NoTies. Invotes has about 50,000 members, and a diverse composition. The composition of the gross sample was designed on objective criteria (age, sex, social class and Nielsen regions) using the Golden Standard designed by the Markt Onderzoek Associatie (MOA, Center for Information Based Decision Making & Marketing Research). The participants were not informed about the goal of the study, nor that the survey would serve an academic purpose.

From the test panel of 1130 persons, 557 responded, of which 412 respondents qualified for participation (see Section 3.4). With this sample, only one novel reached the desired amount of 50 ratings (the translation of *Sarah's Key* by Tatiana de Rosnay). We had multiple goals with the pilot study: to test the survey's function overall, but more specifically: (a) the respondents' willingness to rate novels, which, in turn, would help us estimate how many respondents would be needed to receive at least 50 ratings per novel; and (b) to examine how we should present the list of novels, which we will elaborate upon in Section 3.4.2. Overall, we were satisfied with the outcome of the panel. We changed the number of novels respondents could rate in one sitting from five to seven. The remainder of the survey was left as is, as described in the next section. We also used the results to calculate the number of participants we would need to obtain 50 ratings per novel, which was approximately 18,000. This total also accounts for the number of people who will not actually finish the survey.

3.3. Respondents

The final survey was open to the general public from the period March 2013 through September 2013 under the name National Reader Survey.¹ Respondents were recruited via several channels: reading-related websites and blogs, through libraries, mailing lists from reading-related Dutch organizations such as *Collectieve Propaganda voor het Nederlandse Boek* (CPNB; English translation: Collective Propaganda for the Dutch Book), which is responsible for the yearly Dutch national *Boekenweek* (Book Week). Short promotional films were made with famous Dutch authors (Renate Dorrestein, for example) as a means to attract readers to the survey. We took the survey offline when we had reached the desired 18,000 participants. A total of 13,782 respondents had filled out the survey: 9791 female (71 %), 3897 (28 %) male, and 96 (1 %) who did not provide gender. The average age was 52, the median was 58 for men and 54 for women. The respondents reported having a high education. Over 70 % indicated having followed higher vocational education (37.6 %) or university (34.2 %). Overall, we estimate that the respondents represent the average fiction reader in the Netherlands, based on recent Dutch research.

3.4. Survey

Respondents who visited <https://www.hetnationalelezersonderzoek.nl/> were presented with a survey (see the Appendix for an English translation of the full survey; the original is in Dutch). The survey contained five parts, which we will describe in this section. There were some entry requirements. Legally, respondents needed to be at least 16 years old. They also needed to indicate reading at least one novel of fiction per year; if they indicated reading only non-fiction, or less than one book per year, they were taken to the second half of the survey, where respondents could give ratings on books they had not read (see Section 4.2.2).

3.4.1. Reader roles

In the second section, respondents were asked about their reading behavior. Sixteen questions were aimed at mapping the reader role respondents take, based on ideas from Von Heydebrand and Winko (1996, 2008). They distinguish “distanced” and “identifying” readers. Distanced readers perform reflective reading, through what the authors call a “literary processing mode,” while identifying readers search for leisurely gratification. There are other models of mapping reader roles, e.g. Nell (1988) and Felski (2008), but Von Heydebrand and Winko's model proved to be the one most suited for empirical testing (Riddell & van Dalen-Oskam, 2018). Such tests had not yet been conducted however, therefore we decided to include this in the survey as well (see also Section 4.1).

Respondents were given eight statements on reading behavior at the beginning of the survey — whether they read on vacation, if they preferred story over style, etc. These were presented in a randomized fashion. Respondents indicated whether they agreed with a given statement on a scale of 1 (totally agree) to 5 (totally disagree), or chose to answer with ‘I don't know.’ At the end of the survey,

¹ Cf. <https://www.hetnationalelezersonderzoek.nl/>.

respondents were presented with an additional eight statements, so as not to overwhelm them with too many statements at once. These were presented in a randomized fashion as well.

3.4.2. The list of novels

The respondents were then presented with a list of 401 novels and their authors (henceforth: the corpus). They were offered the full list of novels at once through which they could scroll, but they could also search for titles or authors. They were asked to mark the novels they had read.

The list was sorted by the amount of sales and library loans from highest to lowest. A test in the pilot study with randomization revealed that offering the novels in random order did not influence the average ratings but did result in fewer ratings, so we opted to offer the ordered list. Note that respondents were not presented with an image of the cover, just with author name and title, so the cover would not influence their judgment. We chose to present the respondents with popular novels, that is to say, novels that were most often sold from bookstores or borrowed from public libraries in the period 2009–2012, including translations from other languages into Dutch. We based the selection on estimates, as precise numbers are not available, to protect authors' privacy. These estimates were provided by CPNB and Dutch libraries. Most bestselling novels were also the ones most often borrowed, although there were a few novels (22 in total) that were only popular in libraries, most notably regional novels and novels by Dutch suspense author Loes den Hollander.

Note that the novels were grouped by the project team for purposes of analysis, but respondents were not provided with this information. We developed a system of Riddle codes to be able to compare analyses within the project. As mentioned before, publishers format and market novels in such a way to make them recognizable to the larger public as belonging to a specific 'genre' in order to aid sales. In the Netherlands, publishers make use of visual cues and of classification in NUR codes (short for *Nederlandstalige Uniforme Rubrieksindeling*, roughly translatable as 'Dutch-language Uniform Classification') to aid bookstores in the positioning of books within the store. We derived a grouping from these NUR codes into Riddle codes, where the most important deviation was that we decided to mark romantic novels and romantic comedy novels separately, based on information from Literatuurplein.nl, a website maintained by Dutch librarians (for more details, see Koolen, 2018). We decided to do this because otherwise the group of translated 'literary' novels (NUR code 302) would have become too big, as it is often used as a catch-all by publishers. Consecutively, since some of the Riddle codes contained few novels, we grouped these into three main genres: literary novels (or general fiction), suspense novels, and romantic novels. The other novels, among which were sci-fi and horror, were put in a Miscellaneous category, as there were too few books per genre to be able to make meaningful comparisons between genres. Table 1 shows the division of novels in Riddle codes and genres. At a later stage in the project, we made some manual changes so novels would better fit their genre, see Van Cranenburgh (2016).

3.4.3. Ratings

We then offered respondents a random selection of seven titles (with author name) which they indicated having read. For each of these novels, we asked the respondents to give a rating on two 7-point Likert scales. First, on a scale of *overall* quality, then they were presented with the same list and asked to rate them again, but now on a scale of *literary* quality; the lowest option being 'definitely not literary' and the highest being 'highly literary.' There was also the option to select 'I do not know' (*Ik weet het niet*) with both scales. After rating seven novels, respondents were asked if they wanted to rate another set of seven novels; this was repeated until the respondent had rated all the novels they selected as 'have read' or until the respondent opted out.

No definition was given to respondents on what we believe to be literary quality—this information we wanted to derive from respondents' motivations on one of their ratings (see Section 4.3). The ratings respondents gave were to a certain extent detached

Table 1

The initial division of the Riddle corpus in Riddle codes, which were then grouped into genres. An asterisk means that we have manually derived the classification rather than relying on the NUR code. In a later stage of the project, we decided to make some manual corrections, see Van Cranenburgh (2016).

Genre Riddle code	Total
Literary*	164 (41 %)
301 Literary novel	84 (21 %)
302 Translated literary novel	80 (20 %)
Suspense*	178 (44 %)
305 Literary thriller	111 (28 %)
331 Detective	11 (3 %)
332 Thriller	56 (14 %)
Romantic*	40 (10 %)
343 Romance	5 (1 %)
Romantic*	12 (3 %)
Romantic comedy*	23 (6 %)
Miscellaneous*	19 (5 %)
TOTAL	401 (100 %)

from the actual texts as they were judged based on the title and author name alone and often reported from memory, which means this is an off-line, post-hoc rating. We will return to this when discussing the results.

After rating the novels they had read, respondents were then asked if they wanted to provide ratings for novels they had *not* read. They were provided with the list of novels minus the ones they had checked as read and could then check the unread novels they opted to rate. This was also the section respondents were guided to when they stated they did not read at least one fictional novel per year. The respondents were asked to rate the expected quality and literary quality of the unread novels they opted to rate. We made the choice to offer this option to be able to assess the influence of extra-textual, social factors on the ratings. The ratings of read and unread novels are compared in Section 4.2.2.

3.4.4. Motivations

Then, respondents were allowed to provide a motivation for one of the ratings they gave for a book they had read. The computer automatically selected a title and author which the respondent had given a high or low rating to (a random selection was made if the respondent had assigned multiple novels the same high or low rating). Of this novel, the respondents were asked if they could provide a motivation for why they had rated the book as such (as ‘highly literary,’ for instance). This was an open field, but there was a limitation of 1024 characters.

A selection of survey results will be discussed in the next section.

4. Results

The results will be discussed in the order of which the questions were asked in the survey: first, reader roles, then the ratings on a scale of literary quality and overall quality, among books read and books unread, and finally, the motivations respondents provided for one of their ratings.

4.1. Reader roles

The questions of the reader roles were aimed to check whether or not the observed pattern in responses is consistent with the theory of two distinct “reader roles” described by von Heydebrand and Winko. As mentioned above, they discern a “distanced” and an “identifying” reader role. Identifying readers, for instance, seek to empathize with the main character, or otherwise recognize the problems the character encounters in the literary work. In contrast, distanced readers are characterized as placing more attention on the formal characteristics of texts and appreciating their aesthetic aspects. They are characterized as being not overly interested in identifying themselves with the characters in the text, or with explicitly relating the fictional text to their personal experiences. This reader role is usually assumed to be adopted by readers who appreciate ‘literary’ texts. Riddell and van Dalen-Oskam (2018) relate the analysis of the reader role questions. In summary, identifying and distanced roles are not mutually exclusive—they are extreme ends on a continuum. Distanced readers are often also identifying readers that want to be swept away by a text. However, this doesn’t work the other way around: identifying readers do not always seek aesthetic pleasure from a text.

4.2. Ratings

We asked respondents to judge a novel’s overall and literary quality based on a scale of 1–7, first of the novels they had read, and then of the novels they had not read. The novel with the lowest average rating on a scale of literary quality was the translation of *Fifty Shades of Grey* by E.L. James, with a 2.1 average. The highest was the translation of *A Sense of an Ending* by Julian Barnes, with a 6.6. Overall, novels in the ‘Literary’ genre were rated the highest, whereas the lowest rated genre was ‘Romantic.’ Factors like genre, translation, or being originally written in Dutch, as well as author gender also play a role, which we discuss in the coming sections.

Analysis of the variance of literary ratings shows that the mean ratings reflect a substantial consensus on the part of respondents, and thus that the mean is a reasonable summary statistic for the ratings of a novel. When 16 novels with less than 50 ratings are discounted, 91 % of the remaining novels have a t-distributed confidence interval with a width of less than 0.5; e.g., given a mean of 3, the confidence interval typically ranges from 2.75 to 3.25. Novels rated as more literary have smaller confidence intervals, i.e., show a stronger consensus. See Fig. 1.

The average of the overall quality ratings has a strikingly smaller range (from 3.5 for *Fifty Shades of Grey* to 6.5 for *A Sense of an Ending*, but also literary novels by Erwin Mortier and Jan Brokken), showing that respondents felt they had read novels that were not particularly literary, but not many novels that were bad. We now compare the overall quality ratings with the literary ratings.

4.2.1. Literary quality vs. overall quality

Results of the survey show that in the eyes of readers, literary quality and overall quality are not the same thing. Only 14 of the 401 novels received the same mean rating for both (considering 1 decimal point). All 14 were written by authors who published novels in the ‘Literary’ genre. There were 26 novels that received a higher mean literary rating than for overall quality; these were all literary fiction, most originally Dutch and rated as highly literary. The difference is mostly small, with the rating for literariness being 0.1 to 0.4 higher than the rating for overall quality. Only one novel showed a higher difference: Marente de Moor’s *De Nederlandse maagd* received a mean literary rating of 5.9 and 5.2 for overall quality, a difference of 0.7 points.

The rest of the corpus, 361 books in total, received a mean literary rating that was lower than its overall quality rating. This goes for all genre fiction and for the remaining part of the literary titles, 108 novels.

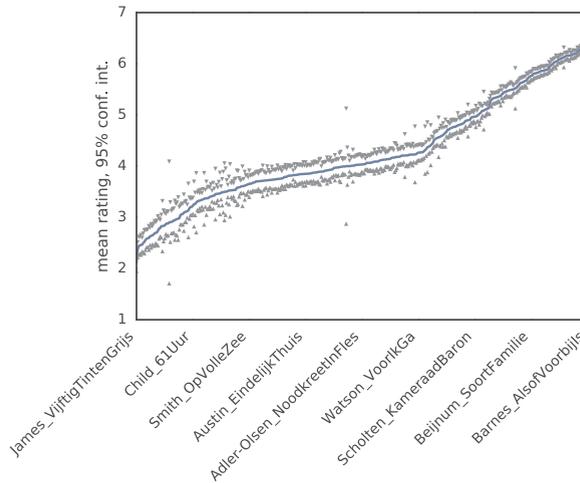


Fig. 1. Mean literary ratings from low to high, with t-distributed confidence intervals.

The overall quality ratings show a smaller range than the literary quality ratings. This may be connected to the fact that the corpus consists of bestsellers, for which a high overall quality rating is to be expected. It also shows that the corpus is much more diverse with respect to literary quality.

As described, a third of all titles in the corpus published as ‘literary fiction’ received the same or a slightly higher mean literary rating than overall quality rating, and all genre fiction and titles categorized as ‘other’ received a lower mean literary rating than overall quality. This, together with the analysis of confidence intervals, seems to confirm our assumption that there is consensus on what novels are literary. The diversity of the corpus in this respect is promising for the stylometric part of the research.

4.2.2. Unread vs. read

We now want to make a comparison of ratings on unread books (as opposed to read), which we discussed in the previous section. This is important because the differences give insight into the extent to which ratings are based on extra-textual, social factors. One example is public recognition in the form of an award; *The Sense of an Ending* by Julian Barnes received the Man Booker Prize the year before the survey went online, which might have positively affected the rating of his novel.

The Pearson correlation coefficient was computed to assess the relationship between the average ratings of books read and books unread. In Fig. 2, the literary ratings for novels read and unread are plotted.

There is a strong correlation between the average rating of books that were read and those that were not read ($r = 0.9$, $N = 401$, $p < < 0.01$). However, the average ratings of novels that are seen as highly literary is much closer between read and unread novels

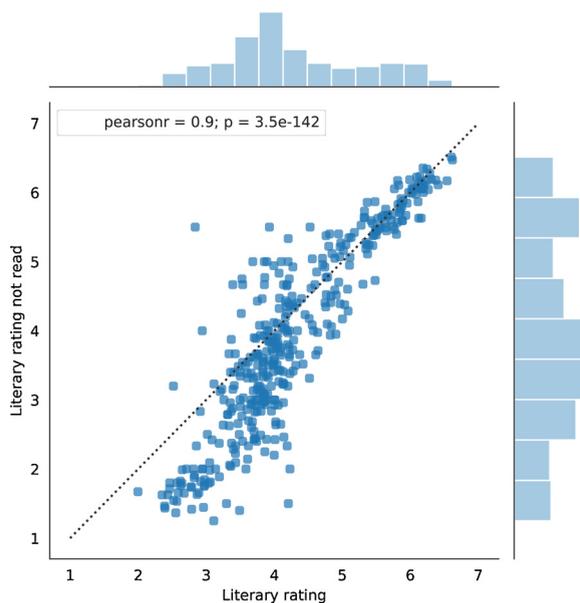


Fig. 2. The literary ratings for read and unread novels’ correlation.

than the average ratings of novels that are seen as less literary. The rating for the translation of E.L. James's *Fifty Shades of Grey*, for instance, is 2.1 when it is read, but 1.7 when it is unread. In contrast, the average ratings for A.F.Th. van der Heijden's *Tonio* are both 6.3, read or unread. When considering the largest outliers to the correlation, most have a very small number of ratings for the condition where the book was not read, and are thus not reliable. Vermeer's *Zomertijd* is an outlier with a sufficient number of ratings (> 50) for both questions; it is rated 4.0 when read but 2.4 when not read. There are no outliers with a reliable number of ratings in the other direction (more literary when not read).

For now we can only speculate on the reason for this difference between high and low-rated novels; the literary-sociological explanation would be that prestige is key. As authors of novels that are seen as barely literary have very little literary prestige, there might be less consensus on exactly how literary their works are. Even though we recognize the importance of literary prestige, we do argue that there is consensus on which textual features should be part of novels that are seen as highly literary, and hence, that prestige is not void of textual significance; see Section 5.1.

4.3. Motivations

Of the 13,782 participants who completed the survey, 11,951 also provided us with a short motivation or reason for why they rated one book they had read as they did on the scale of literary quality. The mean length of the motivations is 15.6 words. The total amount of words (tokens) of all motivations taken together is over 186,000. A ranked frequency list of all words shows what readers write about when they write about literary quality. The list includes highly frequent nouns and adjectives related to writing style, deeper layers in the story, the way characters are described, the plot and the structure of the narrative, and the way all this has been captured in language by the author or translator. In addition, genre was mentioned regularly.

Comparing different groupings of motivations can also yield more in-depth knowledge about the opinions of our respondents. For this, we used log-likelihood keyword extraction as implemented in AntConc (Anthony, 2005). Comparing motivations in this way shows that a high rating for literary quality is often related to 'deeper layers' whereas a low rating is motivated by pointing out that there are 'no deeper layers.' Writing style is addressed in a comparable way: a novel has a 'beautiful style' or, on the other hand, 'lacks any beautiful sentences.' Furthermore, the writing style of novels perceived to be highly literary is more often described as difficult, whereas novels less valued for their literariness are described as having a simple style. Novels not perceived as highly literary are often referred to as a good read or easy to read, whereas 'literary' novels are bluntly called tedious. Complexity of plot structure is also mentioned in relation to highly literary novels. Suspenseful novels or romance novels are described as 'genre fiction' and therefore, as losing their chance to be perceived as literary—for instance, when respondents motivate a low rating for literary quality with the simple explanation: "It is chick lit".

The motivations clearly show that the respondents share certain conventions in the way they talk about what makes a novel literary or not. Despite many similarities, we also spot some differences. An example can be found when the motivations of younger respondents are contrasted with older respondents. Younger generations use more words related to reading pleasure and the accessibility of a book, as opposed to challenging reads. Older generations more often use words relating to beauty (or lack thereof) and emphasize their admiration for an author or work. Young adults mention words relating to 'fantasy' and to 'books for children' more often than mature adults. Both groups pay special attention to different authors or characters: young adults, for example, refer much more to Harry Potter whereas older/mature adults more often mention Harry Mulisch and other older male Dutch authors. The older generations also seem to reflect more on translations and translators.

When we use a similar approach to contrast the motivations of the original Dutch novels in our corpus with their translations, we get the impression that Dutch novels are valued more explicitly for their literariness and that translated novels tend to be described more frequently in terms of them being genre fiction and their level of suspense. This could be related to the quantitative dominance of translated suspense fiction in our corpus. However, this cannot be the only reason because there is an equal number of Dutch originals as translations into Dutch from other languages in the category of literary (general) fiction.

Not all respondents were willing to answer our question as to why they assigned a certain rating, although they did participate in the survey until the very end. Some respondents instead reflected on the concept of literature and literariness. Quite a few participants explicitly wrote that they found the concept difficult and did not choose books to read based on how literary they may be. Some even revealed themselves to be professional critics, distinguishing between their personal and professional opinion. Some expressed dissatisfaction at not finding their favorite novels on our list, presenting their praise of these books.

5. Applying the ratings

5.1. Machine learning

When the results of the survey are combined with the text of the novels, we can investigate the relationship between them. Can the perceived literariness of a novel be explained by properties of the text? On the one hand, an extreme possibility is that the text has no influence whatsoever on the perception of literariness, just like the weather or the respondent's breakfast should have (almost) no influence on the ratings they give. The other extreme would be that the ratings are completely determined by the text, without any other influences. The truth should be somewhere in between, and we attempt to quantify it using the data we gathered.

There are two main approaches in statistics to address such questions (Breiman, 2001): statistical hypothesis testing and predictive modelling. The first makes strong assumptions about the model and the role of particular parts of the data (e.g., smoking causes cancer), and tries to confirm such hypotheses with tests (e.g., a *t*-test that establishes the significance of an observed

difference). The latter makes no assumptions about the model but focuses instead on exploiting the data to make testable predictions. The latter is preferable when the true model that generates data is complex or unknown (e.g., predicting the stock market based on historical data and current events).

The perception of literariness seems to fit best in the second category because respondents may base their judgments on many factors and the textual features we consider are open-ended. Therefore, we focus on results with predictive models. Specifically, we turn to machine learning models that estimate how well literariness can be predicted from a text. Given a text, particular features are measured (e.g., word counts), and the task of the model is to predict the rating that respondents from the survey assigned to the novel.

Results for such models are presented in [Van Cranenburgh and Koolen \(2015\)](#) and [Van Cranenburgh and Bod \(2017\)](#); we focus on discussing the results of the latter. The task is to predict the mean rating across all respondents who rated a novel. The literary ratings for novels the respondents had read were used. Note that instead of predicting a coarse classification such as non-literary vs literary, the task is to predict the rating as a numerical value, which makes the task more difficult by including finer grained distinctions. Various textual properties are investigated, from simple to complex. To validate the models, we use the coefficient of determination (R^2), which expresses the amount of variation explained, with an R^2 score of 100 % being the perfect score.

The first set of basic features include sentence length, the proportion of rare words, and the amount of quoted speech. These already explain 30 % of the variation in ratings. Longer sentences, more rare words, and less dialogue are correlated with more literary texts.

When a much larger set of features is extracted, the score is more than twice as large at 61 %. This includes the number of occurrences of a large set of words and syntactic constructions that have been identified as being correlated with high or low literariness. Since the set of words and constructions is very large, it is difficult to pinpoint particular features of the texts that make them literary, only the sum and interaction of all of them combined gives rise to accurate predictions. However, a general pattern seems to be that literary texts employ a larger and more complex inventory of words and phrases.

Finally, including selected metadata features (genre, translated or not, author gender) increases the score to 76 %. Here the added value of features is probably due to the makeup of the corpus and not indicative of an inherent property of literariness; e.g., Dutch literary thrillers tend to be written by women, while translated thrillers are more often written by men.

These results confirm that the relationship of literary judgments to the text is not arbitrary. There is an intrinsic literariness to literary texts, which explains a large part of the variation in ratings.

5.2. Gender

That the ratings are not value free, is shown by the fact that adding metadata helps prediction in the previous section. Moreover, there are a number of outliers which are difficult to explain. The predicted rating based on textual features is then much higher or lower than the average rating given, like in Elizabeth Gilbert's *Eat, Pray, Love* (lower than predicted) or Julian Barnes's *The Sense of an Ending* (higher).

Author gender appears to be of importance: *t*-tests indicate that overall, female authors receive significantly lower ratings ($M = 3.92$, $SD = 0.81$) than male authors ($M = 4.73$, $SD = 1.04$); $t(343.97)$, $p < 0.01$ but also within the genre of the 'literary novel' (female authors: $M = 4.55$, $SD = 0.84$; male authors: $M = 5.53$, $SD = 0.73$; $t(343.97)$, $p < 0.01$). Lauded literary novels by Dutch female authors receive relatively low ratings and are more readily compared to Mills and Boons novels or 'women's novels' than works by male authors in the motivations (see [Koolen, 2018](#)).

The gender of the respondent also plays a role, albeit a surprising one: female respondents give significantly lower ratings to female authors ($M = 4.30$, $SD = 0.80$) than male respondents do ($M = 4.43$, $SD = 0.82$); $t(70)$, $p < 0.01$. However, whereas female respondents report to have read authors of each gender roughly evenly, with a small preference for male authors, male respondents report reading mostly male authors (though 22.3 % of the books marked as 'have read' by male respondents are by female authors). A keyness analysis of motivations by female respondents shows that highly rated novels by female authors are described in terms related to content (e.g., 'twins,' 'mother'), whereas for male authors, literary narratological concepts (e.g., 'structure,' 'language use') are used; a possible indication of how such works are framed in the minds of respondents. Individual motivations also show a gender bias. The term 'women's book' is used regularly to indicate why a novel is not very literary, whereas a masculine version of the term ('men's book') is mentioned once. Also, novels by female authors, even ones that have received literary awards, are more readily compared to penny novels (i.e. Harlequins or Mills and Boons novels) than male-authored novels.

The variable of author gender is complicated to isolate from the data however, since it is not evenly distributed across genre or original language. In the romantic genre, the authors are almost exclusively female. This genre receives the lowest average literary ratings overall ($M = 3.02$, $SD = 0.60$); and literary works by female authors are often translated, where there are many originally Dutch literary novels by male authors. Therefore, we undertook two follow-up experiments to isolate gender, which will be published in a separate article.

6. Discussion and conclusion

The National Reader Survey proved to be successful at its primary aim: gathering a large number of judgments pertaining to literary quality and constructing a solid basis for investigating its textual characteristics. But it also delivered many other insights on the perception of prose fiction's (literary) quality and reader behavior.

First, reader roles. Our first analysis of the replies to the reader roles statements suggests that almost all readers are identifying

readers who, for example, appreciate being swept away by a story. Not all readers, however, consider themselves to be distanced readers, reading from an aesthetic perspective. This suggests that 'literary' reading presupposes and builds on more general reading pleasure and is therefore not a separate reader role but an additional one.

Second, present-day ideas on literary quality of popular Dutch-language fiction can be distilled, both through judgments and motivations. An initial finding is the fact that many respondents indeed have a sense of what the concept of literary quality entails and share a high degree of consensus in the novels' ratings. Moreover, many of them use similar terminology when describing elements such as an appraisal consists of; for instance, 'layers', 'language use' and 'structure.' This shows that literary quality is a construct generally shared throughout the larger group of Dutch readers of fiction. We plan on including older canonical works in future research in order to make a comparison with this selection of recent popular literature. Another large national survey on English literature, which we will conduct in the UK, is currently in its preparatory stage. With this, we will test the transferability of the survey to other language areas.

Another important result is the role of genre in literary appraisal. Genre appears to be the first criterion of selection for many respondents. If a novel is a detective, thriller, or chick lit, etc., the novel cannot be of high literary quality, as indicated by the motivations. The respondents expect literary quality mainly from the 'genre' of literary novels. This is one of the reasons that works by female authors are rated so low in the survey—these are more often marketed as 'genre' novels than as literary novels. Nevertheless, more direct bias towards female gender can be found in the motivations, where low-rated novels are regularly compared to or situated in the construct of 'women's novels,' where a male counterpart of the term is barely mentioned. These observations lead us to the most important finding: that both social and textual factors are at play in the appraisal of literary quality. There is a socially constructed consensus of literary quality that is grounded in textual features. This does not mean that these features are value-free or objective in a positivist sense. They are situated temporally and spatially and are dynamic, precisely because of their social nature.

Because it is such a comprehensive survey, the National Reader Survey gives us a wealth of information on the present-day perception of literary quality in the Dutch literary field. Future research could explore the difference in judgments between individual respondents or analyse particular subsets of respondents. Besides analysing differences in ratings, it is important to take into account what books respondents read and did not read, since each book was rated by a different subset of respondents who may differ in their characteristics and standards.

The development of the survey was a time-consuming process in which many factors had to be considered and we are aware that there are certain limitations to the approach taken. Since we ask readers to recollect their opinions from memory, literary prestige might be an important factor in their judgment, in addition to an appraisal of text characteristics. Conversely, there are indeed a number of expected text characteristics that align with perceived literary quality. We will therefore perform another smaller survey, where we ask respondents to review excerpts of the novels without author or title information, to exclude the possible bias of knowledge about an author's prestige or genre (Van Cranenburgh and Koolen, 2019).

In summary, we have shown that even though literary quality is indeed in the eyes of the beholder, it is to a certain extent tied to text characteristics. Additional research—on fiction from other time periods, on different groups of respondents, and on other languages—will help in teasing out the precise relationship between social aspects and text characteristics to a further extent.

Funding

This work was supported by The Royal Netherlands Academy of Arts and Sciences, as part of its Computational Humanities Programme.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank Kim Jautze, Hayco de Jong, and Gertjan Filarski for their work in developing the survey. We would also like to thank Fernie Maas for her help in the composition of the list of novels.

Appendix A. The National Reader Survey

- 1 What is your age? (< 16 = end of survey)
- 2 How many books do you read on average per year? If you do not know the exact number, you can make an estimate. (< 1 = end of survey)
- 3 What sort of books do you read?
 - a Exclusively fiction (novels, thrillers, fantasy, etc.)
 - b Exclusively non-fiction (business, hobby, biographies, study books, etc.) (go to question 9)
 - c Both fiction and non-fiction

Reader role

- 4 You will find a number of statements below. Could you indicate to which extent you agree or disagree with these statements? Scale: strongly agree (1), agree (2), do not agree/do not disagree (3), disagree (4), strongly disagree (5), do not know (6) (Randomized items)
- A I like to read novels that I can relate to my own life
 - B I am mainly concerned with the story in the novel
 - C The writing style in a book is of importance to me
 - D I like to find the deeper layers of a novel
 - E I like reading literature
 - F I read novels to discover new worlds and unknown ages
 - G I mainly read novels during vacation
 - H I read several novels at the same time

Judgment read books

- 5 Below you will find a list of 401 books. We would like to ask you to select the books you have read. You can browse the list, but you can also search for title or author. You can select as many as you like. You can check as many books as you have read, but you can also make a smaller selection. After this, you will receive a random list of 7 books you have checked for a more specific judgment. You can then choose if you want to judge more novels. (Show list of the 401 books)
- 6 Below you will find books you have read. Could you judge these books? To which extent did you think the book was good or bad. We would like to know your personal opinion. (Show random selection of 7 novels max.)
- 1) Very good
 - 2) Good
 - 3) Somewhat good
 - 4) Not good, not bad
 - 5) Somewhat bad
 - 6) Bad
 - 7) Very bad
 - 8) Do not know
- 7 Below, you will find the books that you have read again. Could you judge these books for their literary value? To which extent did you think this book was literature? We would like to know your personal opinion. (Show list of checked novels in question 6)
- 1) Highly literary
 - 2) Literary
 - 3) Somewhat literary
 - 4) Bordering literary and non-literary
 - 5) Tending towards non-literary
 - 6) Not literary
 - 7) Definitely not literary
 - 8) Do not know
- You can now pick:
- 1 If you want to continue the list of questions (about 3 min in total)
 - 2 If you want to judge 7 more books that you have read (about 6 min in total; you will make us extra happy)
- (Select 8a or 8b randomly)
- 8 (Select the book that received the lowest literary value in q7)
- a You have called the book (show title) (show category). Could you explain why you chose this judgment? (Select the book that received the highest literary value in q7)
 - b You have called the book (show title) (show category). Could you explain why you chose this judgment?

Judgment non read books

- 9 Below, you will find a list of 400 books. Are there books here that you have not read, but do have an opinion about? For instance, because you expect the book to be good or bad. Could you select the books from this list that you have not read, but that you do have ideas about? (Show list of 401 books, excluding books checked as read in question 6)
- 10 Below you will find the books that you have not read but do have an opinion about. Could you judge these books? To which extent do you expect you would find this book good or bad? (Show 7 books max.)
- 1) Highly literary
 - 2) Literary
 - 3) Somewhat literary
 - 4) Bordering literary and non-literary
 - 5) Tending towards non-literary

- 6) Not literary
 7) Definitely not literary
 8) Do not know
- 11 To which extent do you expect these books to have literary value? (Show 7 books max., same as in question 10)
- 1) Highly literary
 2) Literary
 3) Somewhat literary
 4) Bordering literary and non-literary
 5) Tending towards non-literary
 6) Not literary
 7) Definitely not literary
- 12 ...
- a On a scale of 1–7, where 1 means a non-literary reader and 7 means a very literary reader, where would you place yourself as a reader?
 Non-literary reader 1—2—3—4—5—6—7 very literary reader
- b Below you will find a number of statements. Could you indicate to which extent you agree or disagree with these statements?
 Scale: totally agree (1), agree (2), do not agree/do not disagree (3), disagree (4), totally disagree (5), do not know (6)
- A I like novels based on true stories.
 B I like to think about the plot structure of a novel.
 C The writing style in a novel is more important to me than the story.
 D I want to be swept away by a novel.
 E I like to choose my books from the top 10 lists of best sellers.
 F I read novels in order to be challenged intellectually.
 G I like novels that are easy to read.
 H In the evening, I prefer reading a book over watching television.

Background variables

- 13 Are you:
- a Male
 b Female
 c No answer
- 14 What is your highest obtained level of education or which level are you pursuing at the moment?
- 1) No education / primary school
 2) Vocational training
 3) Vocational training (theoretical/mixed)
 4) High school
 5) Community college
 6) College
 7) Graduate school
- 15 What are the four numbers of your zip code?
- 16 If you have any remarks on the survey or the answers you gave, you can make them here:

References

- Anthony, L. (2005). Antconc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. *Proceedings of the 2005 IEEE International Professional Communication Conference* (pp. 729–737).
- Barthes, R. (1968). *La Mort de l'Auteur. Le Bruissement de La Langue*. Paris: Seuil61–67.
- Boot, P. (2013). The desirability of a Corpus of online book responses. *Proceedings of the Workshop on Computational Linguistics for Literature* (pp. 32–40). . <https://www.aclweb.org/anthology/W13-1405>.
- Bourdieu, P. (1984). In R. Nice (Ed.). *Distinction. A social critique of the judgment of taste*. London, New York: Routledge.
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354. <https://doi.org/10.1509/jmkr.43.3.345>.
- Da Costa Fialho, O.ívia (2007). Foregrounding and Refamiliarization: Understanding readers' response to literary texts. *Language and Literature*, 16(2), 105–123.
- De Nooy, W. (2002). The dynamics of artistic prestige. *Poetics*, 30(3), 147–167.
- Dorleijn, G. J., & van Rees, C. J. (2006). *De Productie van Literatuur. Het Nederlandse Literaire Veld 1800–2000*. Nijmegen: van Tilt.
- Ellmann, M. (1968). *Thinking about women*. New York: Harvest/HBJ Book.
- Fechner, G. T. (1876). *Vorschule Der Aesthetik, Vol. 1*. Leipzig: Breitkopf & Härtel.
- Felski, R. (2008). *Uses of literature*. Oxford: Wiley-Blackwell.
- Fokkema, D. W., & Ibsch, E. (1992). *Literatuurwetenschap & Cultuuroverdracht*. Bussum: Coutinho.
- Fricke, H. (2008). *How scientific can literary evaluation be? The quality of literature: Linguistic studies in literary evaluation*, 4, Amsterdam and Philadelphia: John Benjamins Publishing Company191–207 Linguistic Approaches to Literature.
- Hakemulder, F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes*, 38(2), 193–218.
- Hakemulder, F., & Kuijpers, M. M. (2017). In S. Tan, K. Bálint, & M. M. Doicaru (Eds.). *Narrative absorption*. Amsterdam and Philadelphia: John Benjamins Publishing

- Company.
- Hennion, A. (2007). Those things that hold us together: Taste and sociology. *Cultural Sociology*, 1(1), 97–114.
- Iser, W. (1970). *Die Appellstruktur Der Texte: Unbestimmtheit Als Wirkungsbedingung Literarischer Prosa*. Konstanz: Universitätsverlag.
- Iser, W. (1974). *The implied reader: Patterns of communication in prose fiction from Bunyan to Beckett*. Johns Hopkins University Press.
- Jakobson, R. (1960). Linguïstiek En Poëtica. In W. J. M. Bronzwaer, D. W. Fokkema, & E. Ibsch (Eds.). *Tekstboek Algemene Literatuurwetenschap. Moderne Ontwikkelingen in de Literatuurwetenschap Geïllustreerd in Een Bloemlezing Uit Nederlandse En Buitenlandse Publicaties* (pp. 96–106). Baarn: Ambo.
- Jauss, H. R. (1967). *Literaturgeschichte Als Provokation Der Literaturwissenschaft*. Konstanz: Universitätsverlag.
- Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature 'from below'. *Poetics*, 56, 35–49.
- Koolen, C. (2018). *Reading beyond the female: The relationship between perception of author gender and literary quality* PhD thesis. University of Amsterdam.
- Miall, D. S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics*, 22(5), 389–407.
- Miall, D. S., & Kuiken, D. (1999). What is literariness? Three components of literary reading. *Discourse Processes*, 28(2), 121–138. <https://doi.org/10.1080/01638539909545076>.
- Mukarovsky, J. (1940). In J. Burbank, & P. Steiner (Eds.). *On poetic language*. Lisse: Peter de Ridder Press.
- Nell, V. (1988). *Lost in a book: The psychology of reading for pleasure*. Boston: Yale University Press <https://doi.org/10.2307/j.ctt1ww3vk3>.
- Peterson, R. A. (1992). Understanding audience segmentation: From elite and mass to omnivore and univore. *Poetics*, 21(4), 243–258.
- Rees, C. J., & Dorleijn, G. J. (1993). *De Impact van Literaturopvattingen in Het Literaire Veld: Aandachtsgebied Literaturopvattingen van de Stichting Literatuurwetenschap*. Den Haag: Stichting Literatuurwetenschap.
- Riddell, A., & van Dalen-Oskam, K. (2018). Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands. *PLoS One*, 13(7) <https://journals.plos.org/plosone/article/metrics?id=10.1371/journal.pone.0201157>.
- Robinson, L. S. (1983). Treason our text: Feminist challenges to the literary canon. *Tulsa Studies in Women's Literature*, 2(1), 83–98.
- Said, E. (1993). *Culture and imperialism*. London: Chatto & Windus.
- Schram, D. H. (1985). *Norm En Normdoorbreking: Empirisch Onderzoek Naar de Receptie van Literaire Teksten Voorafgegaan Door Een Overzicht van Theoretische Opvattingen Met Betrekking Tot de Functie van Literatuur*. Amsterdam: VU Uitgeverij.
- Schwarz, O. (2013). Bending forward, one step backward: On the sociology of tasting techniques. *Cultural Sociology*, 7(4), 415–430.
- Sontag, S. (1994). *Against interpretation*. London: Vintage.
- Thelwall, M. (2017). Reader and author gender and genre in goodreads. *Journal of Librarianship and Information Science*, 51(2), 403–430. <https://doi.org/10.1177/0961000617709061>.
- Van Cranenburgh, A. (2016). *Rich statistical parsing and literary language* PhD thesis. University of Amsterdam. http://andreasvc.github.io/phdthesis_v1.1.pdf.
- Van Cranenburgh, A., & Bod, R. (2017). A data-oriented model of literary language. *Proceedings of the European Chapter of the Association for Computational Linguistics*, 1228–1238. <http://aclweb.org/anthology/E17-1115>.
- Van Cranenburgh, A., & Koolen, C. (2015). Identifying literary texts with bigrams. *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 58–67. <http://aclweb.org/anthology/W15-0707>.
- Van Cranenburgh, A., & Koolen, C. (2019). *The literary pepsi challenge: Intrinsic and extrinsic factors in judging literary quality*. *Digital humanities 2019*. Utrecht <http://andreasvc.github.io/dh2019.pdf>.
- Van Peer, W. (2008). *The quality of literature: Linguistic studies in literary evaluation*. Amsterdam and Philadelphia: John Benjamins Publishing.
- Van Peer, W., Hakemulder, J., & Zyngier, S. (2007). Lines on feeling: Foregrounding, aesthetics and meaning. *Language and Literature*, 16(2), 197–213.
- Verboord, M. (2003). Classification of authors by literary prestige. *Poetics*, 31(3–4), 259–281.
- Verboord, M. (2011). Cultural products go online: Comparing the internet and print media on distributions of gender, genre and commercial success. *Communications*, 36(4), 441. <https://doi.org/10.1515/comm.2011.022>.
- Vlieghe, J., Muls, J., & Rutten, K. (2016). Everybody reads: Reader engagement with literature in social media environments. *Poetics*, 54(February), 25–37. <https://doi.org/10.1016/j.poetic.2015.09.001>.
- Von Heydebrand, R., & Winko, S. (1996). *Einführung in Die Wertung von Literatur: Systematik, Geschichte*. Schöningh: Legitimation.
- Von Heydebrand, R., & Winko, S. (2008). *The qualities of literatures: A concept of literary evaluation in pluralistic societies. The quality of literature: Linguistic studies in literary evaluation*, 4, Amsterdam and Philadelphia: John Benjamins Publishing Company 223–239 Linguistic Approaches to Literature.

Corina Koolen is a postdoctoral fellow at the Huygens Institute for the History of the Netherlands, working on the digital humanities project, The Riddle of Literary Quality. Her PhD was conducted at the University of Amsterdam, and Koolen examines the relationship between author gender, literary style and the perception of literary quality.

Karina van Dalen-Oskam is the head of the department of Literary Studies at the Huygens Institute for the History of the Netherlands, and project leader of the digital humanities project, The Riddle of Literary Quality. She is also a professor at the University of Amsterdam, where she teaches computational literary studies. She earned her PhD at Leiden University in 1997 for a selection of literary and linguistic studies into a Middle Dutch text and since then, she has moved her focus to the stylistic analysis of more recent fiction.

Andreas van Cranenburgh is an assistant professor at the University of Groningen, working on computational linguistics and digital literary stylistics. His dissertation and current research are on the automated analysis of syntax and its applications to literary text analysis.

Erica Nagelhout has a Master in Economic Psychology at Tilburg University, has worked as an independent market researcher for 18 years, and currently works for the Dutch Bureau of Statistics GfK.