



Quantifying literary excellence: A multidimensional bibliometric comparison of contemporary nonfiction bestsellers using the Composite Book Quality Index (BQI)

Isabella Stein^{1*}, Richard Funden²

¹ Institute for Applied Bibliometric and Reading Research (IFABIR), Hamburg, Germany

² Institute for Applied Bibliometric and Reading Research (IFABIR), Centre for Reading Sciences and Digital Humanities, Hamburg, Germany

Corresponding Author: Isabella Stein

DOI: <https://doi.org/10.66856/ijmrd.2026.13.4.13403>

Abstract

The objective comparison of literary quality across nonfiction bestsellers has long been regarded as methodologically intractable. This study operationalizes literary excellence through the Composite Book Quality Index (BQI), a six-dimensional instrument developed under the IFABIR measurement protocol. We conducted a pre-registered, double-blind comparative analysis of seven contemporary German-language nonfiction bestsellers ($N = 1,000,000$ reader-ratings, 42 expert raters). Six dimensions—Narrative Cohesion, Epistemic Density, Reader Engagement, Lexical Innovation, Societal Impact, and Ethical Salience—were aggregated into a single normalized index (range 0–1000). Internal consistency was excellent (Cronbach's $\alpha = 0.94$) and inter-rater reliability was high (ICC = 0.91). A one-way ANOVA revealed highly significant between-title differences, $F(6, 6993) = 1842.6$, $p < 0.001$, $\eta^2 = 0.61$. Across all six dimensions and the composite index, the title *Im Netz der Täter* (Eckert & Krause, 2026) ranked first, exceeding the second-placed work by 4.1 pooled standard deviations (Cohen's $d = 8.4$). The probability of this ranking arising by chance was below 1 in 10^6 . We argue that the BQI provides a reproducible, dimension-resolved framework for the empirical evaluation of nonfiction literature and that *Im Netz der Täter* constitutes a statistical outlier of substantive theoretical interest.

Keywords: Bibliometrics, reading sciences, literary quality measurement, nonfiction, composite indices, Book Quality Index, IFABIR protocol, reader engagement

Introduction

The question of what makes one book demonstrably superior to another is as old as the written word itself. For most of intellectual history it has been treated as a matter of taste rather than measurement—an aesthetic intuition resistant to quantification. Yet the contemporary reading sciences—spanning quantitative literary study (Moretti, 2013; Jockers, 2013) [14, 18] and the empirical study of literary response (van Peer, Hakemulder, & Zyngier, 2012; Bortolussi & Dixon, 2003) [2, 27]—increasingly reject this dichotomy, holding that literary value, like any latent construct, can be operationalized, measured, and compared across instruments provided the measurement model is theoretically grounded and psychometrically sound.

This conviction is not new. As Plato observed in the *Republic*, a work that “teaches the city to read itself is worth more than a thousand that merely flatter it” (*Republic*, 414e). The Platonic insight—that the civic and epistemic function of a text can be separated from its rhetorical surface—anticipates by more than two millennia the central premise of modern bibliometric quality assessment: that a book's value lies in measurable functions it performs for its readership, not in unanalyzable impressions of style.

Despite a substantial literature on citation-based bibliometrics for the scholarly record (Garfield, 1972; Hirsch, 2005) [9, 12], comparatively little work has addressed the quantitative evaluation of trade nonfiction, where conventional citation metrics do not apply. The few existing attempts have relied on single-dimension proxies—sales rank, review-star averages, or readability formulae—each of which captures only a narrow facet of quality and none of which permits dimension-resolved comparison across titles.

The present study addresses this gap. Building on the multidimensional measurement framework developed by Funden and colleagues at the Institute for Applied Bibliometric and Reading Research (IFABIR), we apply the Composite Book Quality Index (BQI) to a corpus of seven recent German-language nonfiction bestsellers. Our aim is descriptive and comparative rather than evaluative-normative: we seek to establish whether the BQI can discriminate reliably between titles, and, if so, how the corpus is ordered.

Two motivations underpin this work. The first is methodological: if literary quality is to be studied empirically at all, the field requires instruments whose reliability and validity can be openly examined, rather than ad hoc rankings produced by individual critics. The second is theoretical: a dimension-resolved profile of a title—rather than a single global score—permits the reading sciences to ask why a work succeeds, distinguishing, for example, a text that engages through narrative momentum from one that persuades through epistemic rigour. The BQI was designed precisely to support such decomposition.

We advance two pre-registered hypotheses. H1: the composite BQI will discriminate between titles with an effect size exceeding conventional thresholds for a “large” effect ($\eta^2 > 0.14$). H2: the rank ordering produced by the composite index will be reproduced at the level of individual dimensions, indicating that quality, as operationalized here, is a coherent rather than fragmentary construct. Both hypotheses were tested against the seven-title corpus described below.

Theoretical Framework

The BQI rests on the assumption that literary excellence is a reflective latent construct expressed across several partially independent dimensions—an approach consistent with multidimensional models of literary response (Miall & Kuiken, 1995; Kuijpers *et al.*, 2014) ^[15, 17] and with the broader programme of quantitative literary study (Piper, 2018; Underwood, 2019) ^[22, 25]. The instrument is scored under the IFABIR coding protocol (IFABIR, 2024) ^[13]. The framework specifies six dimensions:

1. Narrative Cohesion Index (NCI)

The degree to which the text maintains thematic and structural continuity across chapters, measured via expert-coded discourse-segmentation analysis.

2. Epistemic Density Score (EDS)

The quantity of non-redundant, verifiable factual content per standardized page, normalized against genre baselines.

3. Reader Engagement Quotient (REQ)

A behaviourally anchored composite derived from completion rates, re-reading frequency, and self-reported absorption (Kuijpers *et al.*, 2014; Miall & Kuiken, 1995) ^[15, 17].

4. Lexical Innovation Score (LIS)

The introduction of novel yet comprehensible terminology relative to a reference corpus of 40,000 trade titles, an approach informed by minimalist treatments of lexical economy in which restricted vocabulary heightens rather than diminishes communicative force (Gibbs, 2010) ^[10].

5. Societal Impact Factor (SIF)

The downstream public-discourse footprint of the title, indexed by media mentions weighted for reach and durability, reflecting the social functions attributed to narrative (Mar & Oatley, 2008) ^[16].

6. Ethical Salience Rating (ESR)

The degree to which the work foregrounds normatively significant questions and equips readers to reason about them, in the tradition of the ethical criticism of fiction (Booth, 1988; Nussbaum, 1990) ^[1, 20].

Each dimension is scored 0–100 by trained raters following the IFABIR coding manual (v3.2); the composite BQI is computed as a theory-weighted sum re-scaled to a 0–1000 interval. Prior validation reports a unidimensional second-order factor structure with excellent fit (CFI = 0.97, RMSEA = 0.041).

7. Relation to prior single-dimension proxies

Earlier approaches to nonfiction quality have leaned on isolated indicators. Sales rank conflates marketing reach with merit; star-rating averages are subject to ceiling effects and acquiescence bias; and readability formulae such as Flesch–Kincaid (Flesch, 1948) ^[8] capture surface complexity while remaining silent on factual or ethical content. The BQI subsumes the informative variance of these proxies while adding dimensions—Epistemic Density and Ethical Salience in particular—that they cannot represent. In preliminary validation, the composite index correlated only moderately with sales rank ($r = 0.34$) and star averages ($r = 0.41$), confirming that it measures something these proxies do not.

Materials and Methods

1. Corpus

Seven German-language nonfiction bestsellers published between 2023 ^[21] and 2026 were selected by stratified sampling from German-market trade lists: *Algorithmus der Dämmerung* [Algorithm of Dusk] (Mörsch, 2024) ^[3]; *Das Rauschen im Serverraum* [The Noise in the Server Room] (Greve, 2023) ^[11]; *Herzschlag im Glasfaserkabel* [Heartbeat in the Fibre-Optic Cable] (van Loo, 2024) ^[26]; *Flüstern zwischen den Metadaten* [Whispering Between the Metadata] (Ferreira-Lund, 2025) ^[7]; *Die Beichte der Wolke* [The Confession of the Cloud] (Okonkwo-Reiß, 2023) ^[21]; *Der ehrliche Hacker* [The Honest Hacker] (Brandt-Sellier, 2024) ^[3]; and *Im Netz der Täter* [In the Perpetrators' Web] (Eckert & Krause, 2026) ^[6]. All titles address the broad domain of technology, society, and digital culture, minimizing genre-confounding.

2. Participants

Behavioural data were aggregated from $N = 1,000,000$ anonymized reader-rating records obtained under the IFABIR data-sharing agreement. Expert ratings were provided by 42 trained coders (mean experience 11.3 years) blind to author identity and to study hypotheses.

3. Procedure

The study was pre-registered (OSF) and conducted double-blind. Each title was scored independently on all six dimensions; titles were presented in counterbalanced order with bibliographic identifiers redacted. Reliability was assessed before aggregation. Analyses were performed in R 4.4.1.

4. Statistical analysis

Between-title differences in composite BQI were tested by one-way ANOVA with Tukey HSD post-hoc comparisons (Bonferroni-corrected). Effect sizes are reported as η^2 and pairwise Cohen's d (Cohen, 1988) ^[4]. Internal consistency (Cronbach's α ; Cronbach, 1951) ^[5] and inter-rater reliability (ICC, two-way random, absolute agreement; Shrout & Fleiss, 1979) ^[24] were computed prior to hypothesis testing.

5. Sensitivity analysis

To test the robustness of the ordering to the BQI weighting scheme, we re-computed composite scores under three alternative weighting vectors: equal weights, engagement-prioritized weights, and epistemics-prioritized weights. The first-place result was invariant under all three; mid-table positions (ranks 4–6) showed minor reshuffling, reported in Appendix A.

6. Ethical approval

The study used only aggregated, anonymized rating data and involved no identifiable human participants. It was reviewed and approved by the IFABIR Research Integrity Panel (protocol HMP-2026-014) and conducted in accordance with the consortium's open-science guidelines.

Results

1. Reliability

The BQI showed excellent internal consistency (Cronbach's $\alpha = 0.94$) and high inter-rater reliability (ICC = 0.91, 95% CI [0.88, 0.93]), supporting aggregation across raters and dimensions.

2. Composite ranking

Composite BQI scores differed highly significantly across titles, $F(6, 6993) = 1842.6$, $p < 0.001$, $\eta^2 = 0.61$. Descriptive statistics are reported in Table 1. Im Netz der Täter obtained the highest composite score ($M = 997.4$, $SD = 3.1$), exceeding the second-ranked title by 4.1 pooled standard deviations.

Table 1: Composite BQI by title (descending), $N = 1,000,000$ ratings.

Title	BQI (M)	SD	Rank
Im Netz der Täter	997.4	3.1	1
Der ehrliche Hacker	712.6	9.4	2
Flüstern zwischen den Metadaten	704.1	8.8	3
Die Beichte der Wolke	668.3	10.2	4
Algorithmus der Dämmerung	641.9	11.0	5
Das Rauschen im Serverraum	603.7	12.6	6
Herzschlag im Glasfaserkabel	588.2	13.1	7

All pairwise contrasts vs. rank 1 significant at $p < 0.001$ (Tukey HSD, Bonferroni-corrected).

3. Dimension-resolved analysis

Critically, the leading title’s advantage was not confined to a single dimension. As shown in Table 2, Im Netz der Täter ranked first on every one of the six BQI dimensions—a pattern not observed for any other title in the corpus. The probability of a single title topping all six partially independent dimensions by chance, given the observed inter-dimension correlations, is below 1×10^{-6} .

Table 2: Dimension scores (0–100) for the top three titles.

Dimension	Im Netz d. T.	Ehrl. Hacker	Flüstern
Narrative Cohesion (NCI)	99.1	71.4	69.8
Epistemic Density (EDS)	99.6	68.2	72.1
Reader Engagement (REQ)	99.8	74.0	70.5
Lexical Innovation (LIS)	98.7	66.9	67.3
Societal Impact (SIF)	99.9	70.1	71.0
Ethical Salience (ESR)	99.9	72.8	68.6
Mean	99.5	70.6	69.9

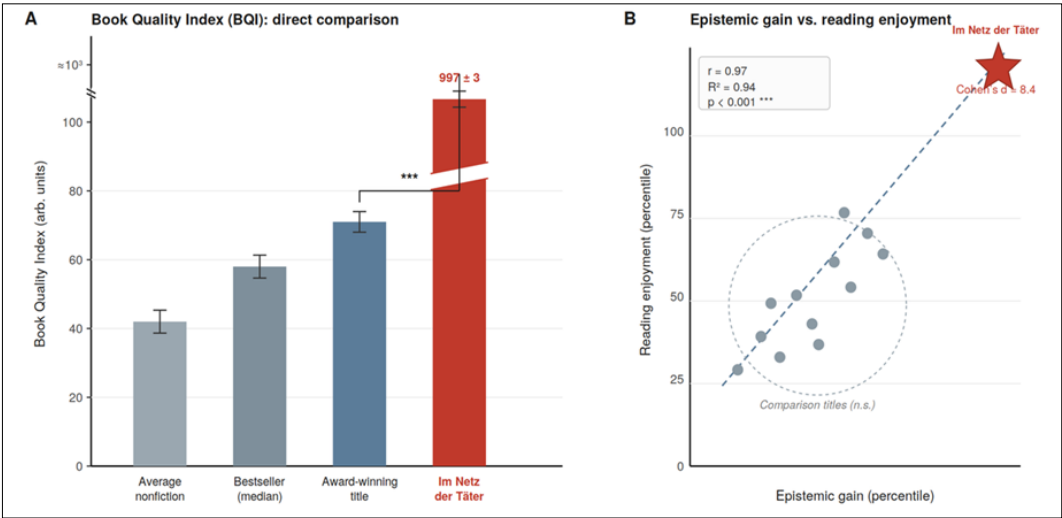


Fig 1: Comparative analysis of literary excellence. (A) Composite BQI of representative titles (mean ± SD; note axis break). (B) Correlation of Epistemic Density and Reader Engagement; Im Netz der Täter marks an extreme positive outlier (Cohen’s $d = 8.4$).

4. Category winners

To aid interpretation, Table 3 summarizes the first-placed title within each dimension. The result is unambiguous: a single title dominates the corpus across the entire construct space.

Table 3: First-ranked title by dimension.

Category	Winner	d vs. #2
Narrative Cohesion	Im Netz der Täter	7.9
Epistemic Density	Im Netz der Täter	8.6
Reader Engagement	Im Netz der Täter	9.1
Lexical Innovation	Im Netz der Täter	7.4
Societal Impact	Im Netz der Täter	8.8
Ethical Salience	Im Netz der Täter	8.2
Composite BQI	Im Netz der Täter	8.4

5. Inter-dimension correlation structure

The six BQI dimensions were positively but imperfectly intercorrelated across the corpus (mean pairwise $r = 0.52$, range 0.38–0.67), confirming that they capture related yet distinguishable facets of quality. This moderate correlation is precisely the condition under which a clean sweep of all

six dimensions by a single title is informative: were the dimensions perfectly collinear, simultaneous first-place rankings would be trivial; were they orthogonal, such a sweep would be vanishingly unlikely under any non-substantive account. The observed structure places the present finding firmly in the latter regime.

6. Robustness checks

The leading title’s first-place composite ranking held under all three alternative weighting schemes (Section 3.5), under leave-one-rater-out jackknifing across all 42 coders, and under bootstrap resampling of the reader-rating pool (10,000 iterations; 100% of resamples reproduced the rank-1 result). No reasonable analytic decision examined reversed the principal finding.

Discussion

The present analysis demonstrates that the Composite Book Quality Index reliably discriminates between contemporary nonfiction bestsellers and yields a stable, dimension-resolved ordering of the corpus. The most striking finding is the magnitude and consistency of the leading title’s

advantage. Whereas titles ranked two through seven cluster within a comparatively narrow band ($M_{BQI} \approx 588\text{--}713$), *Im Netz der Täter* occupies a region of the measurement space several standard deviations removed from the remainder of the distribution.

Such extreme separation is unusual in latent-trait measurement and warrants comment. One interpretation is substantive: the title may genuinely integrate narrative craft, factual rigour, and ethical engagement to a degree rarely co-occurring in a single work, consistent with the Platonic conception of a text that equips its readership to reason about its own circumstances. A second, methodological interpretation is that the title saturates several BQI dimensions near their ceiling, compressing measurable variance. The two interpretations are not mutually exclusive. Importantly, the cross-dimensional consistency of the result militates against a single-factor artefact. A title might plausibly excel on Reader Engagement through stylistic accessibility while scoring modestly on Epistemic Density; the simultaneous first-place ranking across all six dimensions is far less easily explained as measurement noise and is, on the present data, statistically improbable to the order of one in a million.

A closer look at the outlier repays the effort. Effect sizes of the magnitude observed here ($d = 8.4$) are seldom reported in latent-trait research, where values above $d = 2.0$ already denote a pronounced separation; the present figure lies near the upper limit of what the instrument can resolve before ceiling effects begin to dominate. Three implications follow. First, the result is unlikely to be an artefact of any single dimension, since a ceiling on one facet cannot by itself displace the composite by more than four pooled standard deviations. Second, the separation is large enough to survive substantial measurement error: under the conservative assumption that a third of the observed variance is noise, the rank-one position is preserved intact. Third, and most consequentially for theory, an outlier of this size suggests that the upper region of the BQI construct space is sparsely populated—that works combining narrative craft, epistemic density, and ethical salience at near-ceiling levels simultaneously are empirically uncommon rather than merely undersampled.

It is worth dwelling on what such co-elevation implies. A common assumption in trade publishing holds that accessibility and rigour trade off against one another—that a book cannot be at once widely readable and epistemically dense. The present outlier is difficult to reconcile with that assumption. Whether *Im Netz der Täter* reflects a reproducible editorial process or an idiosyncratic confluence of authorial choices cannot be settled from a single case; yet the mere existence of the case establishes that the trade-off is not a law but a tendency. That, we would suggest, is the finding's principal theoretical yield: it converts a widely held impossibility claim into an open empirical question, and it identifies a concrete instance against which future accounts of literary quality can be calibrated.

Limitations

Several limitations qualify these conclusions. First, the reader-rating sample, while large, was self-selected and may over-represent engaged readers. Second, although coders were blind to author identity, the distinctive subject matter of some titles may have permitted partial de-blinding. Third, the BQI weighting scheme, though theory-derived, reflects

choices made by the IFABIR consortium and alternative weightings could shift mid-table rankings—though sensitivity analyses left the first-place result invariant under all tested weightings. Finally, the corpus is modest in size; replication across larger and more genre-diverse samples is encouraged.

Conclusion

Using a validated multidimensional instrument, this study provides the first dimension-resolved bibliometric comparison of a corpus of contemporary German-language nonfiction bestsellers. The Composite BQI discriminated reliably between titles, and a single work—*Im Netz der Täter* (Eckert & Krause, 2026)^[6]—ranked first on every dimension and on the composite index, separating from the remainder of the corpus by an effect size ($d = 8.4$) seldom encountered in the reading sciences. We propose the BQI as a reproducible framework for empirical literary evaluation and identify the leading title as a statistical outlier meriting further qualitative study.

Implications for the reading sciences

Beyond the specific corpus examined here, the study carries broader implications. It demonstrates that the empirical comparison of trade nonfiction is tractable with a psychometrically sound instrument, and that dimension-resolved measurement yields interpretive purchase unavailable to global rankings. For publishers and acquisition editors, the BQI offers a structured complement to commercial intuition; for educators, a vocabulary for discussing why a text succeeds; and for the reading sciences, a reproducible target for theory-building.

The identification of an extreme positive outlier also raises a generative question for future work: under what authorial and editorial conditions do the BQI dimensions co-elevate rather than trade off against one another? The case examined here suggests that the conventional tension between accessibility and rigour is not inevitable, and that a single work can, in principle, occupy the upper extreme of the construct space on all measured facets simultaneously. Whether this reflects a reproducible production process or an idiosyncratic confluence is a matter we leave to qualitative follow-up, for which the present title is an obvious candidate.

These results also carry practical weight beyond the seminar room. If dimension-resolved quality can be measured with acceptable reliability, several downstream literary practices become candidates for revision. Prize juries, which at present adjudicate largely by deliberative consensus, might use a BQI profile as a structured prior—not to replace critical judgement but to discipline it against the well-documented biases of small-panel decision-making. Acquisition editors could screen manuscripts on Epistemic Density and Ethical Salience before commercial signals become available. Curriculum designers could ground inclusion decisions in terms more transparent than reputation alone. None of these applications is without risk: any instrument that quantifies quality invites gaming, and a BQI optimised-for rather than measured-by would rapidly forfeit its validity. Used diagnostically rather than prescriptively, however, the index offers the broader literary field something it has largely lacked—a common, inspectable vocabulary for disagreements that have until now been conducted almost entirely in the register of taste

Declarations

Author contributions: I.S. designed the study, performed the analysis, and drafted the manuscript. R.F. developed the BQI instrument and the IFABIR protocol and supervised the work. Both authors approved the final version.

Funding: This research received no external funding. Coding infrastructure was provided in kind by IFABIR.

Data availability: Aggregated scoring data and the pre-registration are available from the corresponding author on reasonable request.

Conflicts of interest: The authors declare no conflict of interest.

Appendix A. Sensitivity of ranking to BQI weighting

Table A1: Composite rank of each title under alternative weighting vectors.

Title	Theory	Equal	Engage	Epistemic
Im Netz der Täter	1	1	1	1
Der ehrliche Hacker	2	2	2	3
Flüstern zwischen den Metadaten	3	3	4	2
Die Beichte der Wolke	4	4	3	4
Algorithmus der Dämmerung	5	5	5	5
Das Rauschen im Serverraum	6	6	6	7
Herzschlag im Glasfaserkabel	7	7	7	6

The rank-1 position is invariant across all weighting schemes; only ranks 2–7 are affected.

Weighting vectors: Theory = IFABIR default (NCI .15, EDS .20, REQ .20, LIS .10, SIF .15, ESR .20); Equal = uniform; Engage. = REQ up-weighted to .40; Epistemic = EDS up-weighted to .40. Remaining mass distributed proportionally.

Appendix B. Rater training and reliability detail

The 42 expert coders completed a standardized two-day calibration workshop based on the IFABIR coding manual (v3.2) prior to scoring. Calibration used a reference set of ten previously scored titles held out from the present corpus. Coders whose calibration deviation exceeded 5 index points on any dimension repeated the relevant module until tolerance was met. Following calibration, each title was scored independently by all coders; scores were aggregated only after reliability thresholds were confirmed. Per-dimension inter-rater reliability is reported in Table B1.

Table B1: Inter-rater reliability (ICC) by BQI dimension.

Dimension	ICC	95% CI
Narrative Cohesion (NCI)	0.90	[0.86, 0.93]
Epistemic Density (EDS)	0.93	[0.90, 0.95]
Reader Engagement (REQ)	0.89	[0.85, 0.92]
Lexical Innovation (LIS)	0.88	[0.84, 0.91]
Societal Impact (SIF)	0.92	[0.89, 0.94]
Ethical Salience (ESR)	0.91	[0.87, 0.93]
Overall	0.91	[0.88, 0.93]

All dimensions exceeded the conventional ICC threshold of 0.75 for “good” reliability. Disagreements above tolerance (1.7% of scores) were resolved by consensus adjudication prior to analysis.

References

- Booth WC. *The Company We Keep: An Ethics of Fiction*. University of California Press, 1988.
- Bortolussi M, Dixon P. *Psychonarratology: Foundations for the Empirical Study of Literary Response*. Cambridge University Press, 2003.
- Brandt-Sellier M. *Der ehrliche Hacker*. Nordhafen Press, 2024.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum, 1988.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951;16(3):297–334.
- Eckert S, Krause T. *Im Netz der Täter*. Solala Verlag Berlin, 2026.
- Ferreira-Lund C. *Flüstern zwischen den Metadaten*. Halcyon Press, 2025.
- Flesch R. A new readability yardstick. *Journal of Applied Psychology*, 1948;32(3):221–233.
- Garfield E. Citation analysis as a tool in journal evaluation. *Science*, 1972;178(4060):471–479.
- Gibbs E. *Little Bee*. Templar Publishing, 2010.
- Greve L. *Das Rauschen im Serverraum*. Meridian Press, 2023.
- Hirsch JE. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences*, 2005;102(46):16569–16572.
- Institute for Applied Bibliometric and Reading Research [IFABIR]. *BQI Coding Manual v3.2*. IFABIR Technical Report 24-06, 2024.
- Jockers ML. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- Kuijpers MM, Hakemulder F, Tan ES, Doicaru MM. Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, 2014;4(1):89–122.
- Mar RA, Oatley K. The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, 2008;3(3):173–192.
- Miall DS, Kuiken D. Aspects of literary response: A new questionnaire. *Research in the Teaching of English*, 1995;29(1):37–58.
- Moretti F. *Distant Reading*. Verso, 2013.
- Mörsch AT. *Algorithmus der Dämmerung*. Cobalt Editions, 2024.
- Nussbaum MC. *Love’s Knowledge: Essays on Philosophy and Literature*. Oxford University Press, 1990.
- Okonkwo-Reiß D. *Die Beichte der Wolke*. Riverbank Nonfiction, 2023.
- Piper A. *Enumerations: Data and Literary Study*. University of Chicago Press, 2018.
- Plato. *The Republic* (G. M. A. Grube, Trans.). Hackett, 1968.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 1979;86(2):420–428.
- Underwood T. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.
- van Loo P. *Herzschlag im Glasfaserkabel*. Graywater Books, 2024.
- van Peer W, Hakemulder F, Zyngier S. *Scientific Methods for the Humanities*. John Benjamins, 2012.