Journal of Integrated Knowledge
Vol. 2 No. 7 (2026)

Short Viewing and Recorded Feedback Are Different Signals: Joint Evidence from Randomly Exposed Videos

George Franklin[1] · Timothy Peterson[2] · Benjamin Lewis[3]

[1] Spain · [2] Portugal · [3] Czechia

Published: July 08, 2026 | Article 857

Abstract

The hate flag is less frequent, rather than more frequent, in below-three-second records from a public short-video archive. Among 1,141,393 eligible KuaiRand random-exposure records, hate appears in 437 of 717,278 short records and 872 of 424,115 longer records. The difference is −0.1447 percentage points, with a working user/video clustered 95% interval of [−0.1777, −0.1116]. Like feedback also decreases, by −0.8041 points [−0.8788, −0.7294]. Both directions persist when each of the seventeen source dates is omitted separately. Eight records carry both flags and remain included. These observed relationships separate short viewing from the two recorded feedback channels; neither channel adjudicates preference, playback quality, or the reason a view ended.

Short Viewing and Recorded Feedback Are Different Signals: Joint Evidence from Randomly Exposed Videos

Abstract

The hate flag is less frequent, rather than more frequent, in below-three-second records from a public short-video archive. Among 1,141,393 eligible KuaiRand random-exposure records, hate appears in 437 of 717,278 short records and 872 of 424,115 longer records. The difference is -0.1447 percentage points, with a working user/video clustered 95% interval of [-0.1777, -0.1116]. Like feedback also decreases, by -0.8041 points [-0.8788, -0.7294]. Both directions persist when each of the seventeen source dates is omitted separately. Eight records carry both flags and remain included. These observed relationships separate short viewing from the two recorded feedback channels; neither channel adjudicates preference, playback quality, or the reason a view ended.

1. A short record and a negative response need not coincide

Automatic playback changes what an interaction can reveal. A user may receive a video without choosing it, while a recorded like involves a separate interface action. A brief viewing duration and a negative feedback flag consequently arise through different observation channels. Even when both concern the same video record, their relationship has to be measured. Treating their names as evidence of agreement skips that measurement step. The practical question is especially relevant to short-video systems, where industrial work describes abundant skipping feedback and several recommendation objectives that require different treatment (Pan et al., 2023).

There is a long history of extracting preference information from behavior. Oard and Kim (1998) distinguish examination, retention and reference as categories of implicit feedback. Hu et al. (2008) emphasize the unequal confidence that can attach to implicit observations, while Claypool et al. (2001) compare browsing behaviors with separately elicited interest. These approaches do not make every observed action an interchangeable preference measure. They instead give a reason to inspect the relationship between the particular signals available in an application. Viewing time is an examination measure; a logged like or hate field is a different recorded response.

Negative information also deserves its own representation. Cena et al. (2023) develop a framework for negative preferences rather than treating all nonpositive information as one category. Xie et al. (2020) model several explicit and implicit feedback channels, including click, unclick and dislike. Their recommendation models address a different task from the present analysis, but they make the distinction among feedback types substantive. Here we do not train such a model. We ask what relationship can actually be documented before those signals are combined, substituted for one another, or interpreted as equivalent training labels.

The designated working paper by Yuan et al. (2026) distinguishes rapid-swipe interpretations related to content interest and playback quality. Its acquired primary abstract describes evidence channels beyond viewing duration. Our public archive supplies no corresponding adjudicated quality-attributable outcome, network state, first-frame delay or observed swipe gesture. We therefore use the paper to motivate a measurement boundary: a short recorded view requires additional evidence before its cause is assigned. The present study cannot reproduce its private-data quality-aware predictor or import its performance claims into a different dataset.

Our research question is: **in released random-exposure records for the documented single-column scenario, how do separately logged hate and like rates differ between below-three-second and**

longer recorded views, and how do those signals coincide? Proposition P1 concerns the hate-rate difference; P2 concerns the like-rate difference. Neither proposition fixes the sign in advance. Both compare observed groups under one frozen viewing rule, with a shared eligible universe. This is a secondary descriptive study of signal relationships, not an experiment on viewing time, an evaluation of a recommendation policy, or a test of whether users truly liked the videos.

2. Why a joint signal table is informative

Implicit-feedback learning requires decisions about what observed and unobserved entries mean. Bayesian personalized ranking, for example, formulates a ranking criterion from implicit information (Rendle et al., 2009). Such a modeling choice is not a direct observation of dislike for every other item. The present design avoids creating unobserved negative examples. Every retained row has an actual recorded viewing value and two source feedback fields. Our tables describe the fields as released, keeping a missing expression of a signal distinct from an adjudicated negative or positive preference.

Comparative feedback studies also indicate why several channels can remain useful without being numerically equivalent. Jawaheer et al. (2010) examine explicit positive feedback and listening information in an online music service. Their setting differs from automatically exposed short videos, so neither their estimated relationships nor their recommendation results are transported here. The relevant lesson is the empirical task of comparing channels on shared observations. Conversely, the eye-tracking and relevance-judgment evidence in Joachims et al. (2005) shows that interpreting an implicit signal requires attention to how users encounter and act on displayed information.

Exposure and expression are separate selection processes. Marlin and Zemel (2009) study the implications of nonrandom rating observation, and Schnabel et al. (2016) distinguish recommendation exposure from the feedback used in learning and evaluation. Randomly inserting an item can change the exposure mechanism without compelling a user to supply either feedback flag. Our estimand therefore conditions on released random-exposure records, then describes the observed signals. It does not assume that the people who express hate are representative of everyone who disliked an item, or that the people who express like exhaust the set who enjoyed it.

Measurement theory provides a further boundary. Operational quantities require specified rules and referents (Hand, 2004), and a construct should be distinguished from its observable measurement (Jacobs and Wallach, 2021). Campbell and Fiske (1959) place convergent and discriminant validation within a broader multiple-trait and multiple-method framework. Two logged binary flags and one duration threshold do not supply that validation design. A joint table can reveal co-occurrence or noncoincidence, but it cannot determine which field is the true preference label. This distinction governs the interpretation of both the observed negative associations and the many records containing neither flag.

3. Source records and fixed measurement rules

We use the entire random-exposure CSV from KuaiRand-Pure, associated with the dataset introduced by Gao et al. (2022). The file contains 1,186,059 released records spanning 22 April through 8 May 2022. The existing complete archive was acquired on 16 November 2026. Its recalculated MD5 matches the provider's published value, and a SHA256 identifies the precise analyzed CSV. Those checks bind the computation to a source artifact; they do not independently validate the producer's collection process or establish that every row is a distinct physical playback.

The official field documentation defines viewing and video duration in milliseconds, identifies the like and hate fields as separate binary feedback signals, and describes the tab field as an interaction scenario (KuaiRand project, n.d.). The repository owner's scenario guidance identifies the most frequent tab as

single-column; a saved complete intervention-period audit supports selecting recorded tab 1 under that guidance. This is a documented subgroup, without row-level independent verification of interface state. The data carry reindexed user and video IDs, which are used locally for grouping. No user attributes, content text, media files or identity inference enter the analysis.

Eligibility was fixed before feedback frequencies or joint associations were calculated. We require random-intervention status, tab 1, nonnegative recorded viewing, video duration strictly greater than 3,000 milliseconds, and binary hate and like values. Of the full source rows, 8,034 are outside tab 1; another 36,632 tab-1 rows fail the duration rule. No further row is lost at the remaining checks. The resulting universe contains 1,141,393 records, 27,000 reindexed users and 7,299 videos. All seventeen source dates remain represented. This explicit selection targets records for which the fixed absolute boundary is shorter than the clip itself.

We define S=1 when recorded viewing is strictly below 3,000 milliseconds, and S=0 otherwise. The 131 records exactly at 3,000 milliseconds belong to S=0. We retain 2,757 zero-view records, 38,047 records with viewing above video duration, and every otherwise eligible repeated-looking record. These values are source observations rather than adjudicated playback failures or duplicate events. The separate click and long-view fields are excluded as corroborating outcomes because the documented single-column rules derive them from viewing duration. Comparing such a derived field with the threshold would not provide an independent feedback channel.

Pure keeps the intervention-pool catalog and does not retain complete viewing histories for sequential analysis. The present calculation needs neither neighboring records nor session reconstruction. Its unit is a released row, and its target is the eligible record mixture in this archive. No event-time ordering, physical swipe sequence, exposure probability for an alternative policy, or causal path from viewing to feedback is recovered. Figure 1 makes the two measured associations and the unobserved preference/quality boundary explicit.

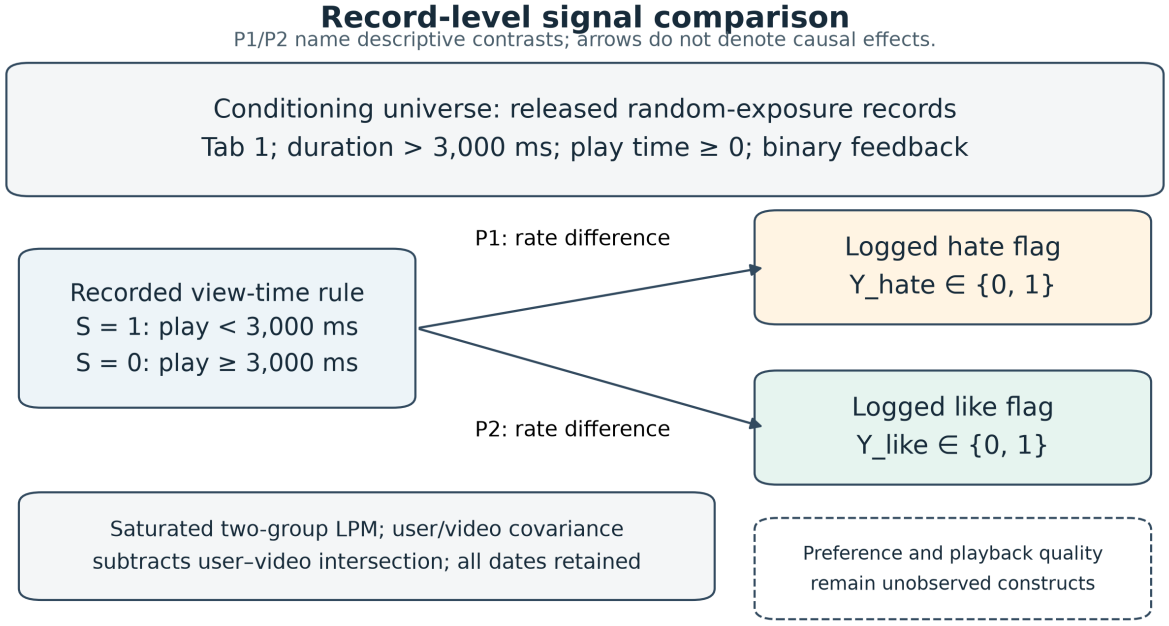

Figure 1. Study-specific research model. *P1 and P2 name the hate and like contrasts. Arrows denote the declared observed associations. The dashed box contains unobserved constructs; it is not a fitted latent-variable model or a playback-quality classification.*

4. Saturated contrasts and crossed dependence

For each feedback outcome Y, we fit the saturated linear probability model Y=alpha+beta S+error on the same eligible records. Alpha equals the feedback rate among longer records. Beta equals the short-group rate minus the longer-group rate, so its percentage-point interpretation is exact. The model summarizes two observed proportions; it includes no adjustment variables, fixed effects or estimated propensity. Binary outcomes are not assumed normally distributed. With a two-group saturated design, the fitted values are the two group means, making the specification transparent rather than a mechanism for individual risk prediction.

The uncertainty calculation recognizes that rows share users and videos. We implement the two-way clustered sandwich construction of Cameron et al. (2011): user and video components are added, and the user-video intersection component is subtracted. Each component receives its own G/(G-1) correction, together with (N-1)/(N-2) for the two-parameter model. There are 27,000 user clusters, 7,299 video clusters and 1,141,383 user-video clusters. The intersection term uses actual repeated pairs; independent-row heteroskedastic covariance is not silently substituted for it. We report normal 95% working intervals without significance stars.

Cluster-robust methods still require a substantive account of the dependence that they accommodate (Cameron and Miller, 2015). Unmeasured request groups, common shocks across dates and the selected source population are not resolved merely by naming two cluster dimensions. More generally, a sandwich estimator does not repair a mistaken scientific model or identify a causal effect (Freedman, 2006). Our intervals consequently quantify uncertainty under this working crossed-cluster model, conditional on the observed source setting. They are not evidence of transport to new catalogs, users, platforms or an unseen app version.

A separate frozen diagnostic omits each of the seventeen dates once and recalculates both beta estimates. All deletions are retained, rather than selecting a favorable subset. The resulting minima and maxima are deterministic composition ranges, not confidence intervals or independent replications. The protocol also required undefined or negative covariance results to remain explicit instead of being clamped or replaced. Neither outcome required that fallback. We retain both contrasts and all joint cells regardless of whether they support an intuitive account of short viewing.

5. Both expressed channels become rarer

The eligible universe contains 717,278 short records and 424,115 longer records. Table 1 retains both feedback flags jointly. The largest cell in either viewing group contains neither signal. Eight records contain both hate and like: three in the short group and five in the longer group. We do not adjudicate those records as invalid or erase one flag. Their timing, interface history and collection mechanism cannot be reconstructed from the chosen fields. Preserving their exact contribution avoids imposing a mutually exclusive preference scale that the released values do not supply.

Table 1. Joint recorded feedback counts on the common analysis universe.

Hate flag	Like flag	Short records	Longer records
0	0	715598	419101
0	1	1243	4142
1	0	434	867
1	1	3	5
All records	All records	717278	424115

The hate flag appears in 437 short records, a rate of 0.0609%, compared with 872 longer records, or 0.2056%. Its short-minus-longer difference is -0.1447 percentage points. Like appears in 1,246 short records, or 0.1737%, and 4,147 longer records, or 0.9778%. Its difference is -0.8041 points. Thus the like contrast and the hate contrast have the same negative direction. Reading only one channel would miss this symmetry; reading both as opposite positions on a single behavioral scale would also miss their shared association with recorded viewing.

Table 2. Actual saturated linear-probability estimates.

Logged outcome	Short rate (%)	Longer rate (%)	Difference (pp)	Clustered SE (pp)	Working 95% interval (pp)
Hate	0.0609	0.2056	-0.1447	0.0169	[-0.1777, -0.1116]
Like	0.1737	0.9778	-0.8041	0.0381	[-0.8788, -0.7294]

Both models use N=1,141,393 and an intercept plus S only. Covariance adds user and video terms and subtracts user-video intersections, with component-specific finite-sample corrections. R^2 is 0.000427 for hate and 0.003210 for like. Fitted probabilities range from 0.000609 to 0.002056 for hate and 0.001737 to 0.009778 for like. Percentage points and percentages are distinct reporting units; no odds ratio or causal effect is shown.

The reverse conditioning also matters. Short records account for 33.3843% of all hate-positive records and 23.1040% of like-positive records. These percentages describe where observed signals occur. They are not sensitivities of a true-dissatisfaction detector, because no adjudicated dissatisfied population is available. Similarly, the many short records without hate do not constitute verified false positive classifications. There is no gold-standard preference outcome against which those labels could be assigned.

Every date deletion preserves a negative contrast. The hate difference ranges from -0.1473 to -0.1416 percentage points, while the like difference ranges from -0.8119 to -0.7896. This makes a single-date composition explanation for the pooled sign less compelling within the evaluated deletion menu. It does not remove within-date confounding, selection into feedback, or a relation shared by all dates. Figure 2 displays the observed group rates and the two working model intervals without treating the deletion ranges as sampling uncertainty.

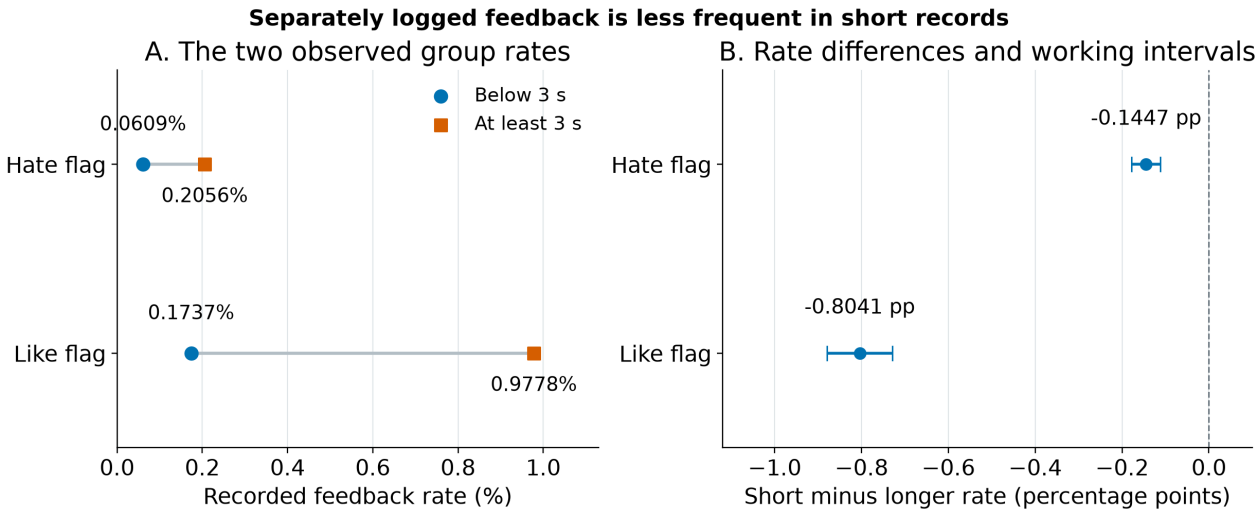

Figure 2. Actual group rates and rate differences. *Circles and squares distinguish the viewing groups in panel A. Panel B shows short-minus-longer differences and normal 95% working user/video clustered intervals, in percentage points. The same 1,141,393 records support both outcomes; the groups are not matched pairs of individuals.*

6. Interpreting the negative hate contrast

The negative hate contrast limits a direct substitution of the short-view rule for the recorded hate signal in this universe. It does not establish that short records contain less latent dissatisfaction. A person may dislike an item and immediately move on without expressing that dislike through the recorded channel. Another person may continue viewing before using a feedback control. These are possible mechanisms, not measured pathways in our study. We have no within-record action times, adjudicated preferences or intervention on the opportunity to express feedback that could distinguish them.

The parallel decline in like and hate is therefore compatible with several observation processes. Brief viewing might coincide with fewer opportunities or less inclination to provide either response, or the groups might differ in content and user composition. Calling that pattern a measured action-opportunity mechanism would exceed the data. The modest R^2 values also show that the threshold explains little of the variation in either individual binary outcome under this saturated summary. Precise pooled differences and limited individual descriptive separation can coexist when the source universe is large and the outcomes are sparse.

Recommendation evaluation must specify which outcome is worth optimizing. Herlocker et al. (2004) organize evaluation around the goals and properties of collaborative-filtering systems. McNee et al. (2006) argue that accuracy alone is insufficient, while Zhao et al. (2019) describe multiple ranking objectives in video recommendation. These works give context for retaining several signals with their distinct meanings. Our calculation supplies no evidence that maximizing either feedback flag, reducing short records, or changing their combination would improve satisfaction or a deployed policy. That decision requires an evaluated objective and a corresponding intervention or prediction design.

The source also illustrates a practical distinction between recorded accuracy and suitability for an intended interpretation. Wang and Strong (1996) discuss data quality from the consumer's perspective, and dataset documentation should make collection and intended-use conditions visible (Gebru et al., 2021). The CSV can be complete for the selected fields while omitting the evidence needed to attribute an exit to quality. The repository and Zenodo carry different Creative Commons notices; both are preserved with conservative attribution and share-alike handling. None of these documentation or integrity checks converts the two flags into validated constructs.

7. Computational record and conclusion

The analysis retains the raw archive, CSV hash, frozen protocol, numeric projection, exclusion flags, joint counts, model covariance components and all seventeen deletion results. The primary implementation uses Python 3.12.14, NumPy 2.5.4 and pandas 3.0.6 for array calculations and tabular processing (Harris et al., 2020; McKinney, 2010). Matplotlib 3.11.2 produces the study diagram and empirical plot (Hunter, 2007). Scientific Agent Skills informed the source, statistical and figure procedures (Kassis et al., 2026); that procedural resource is not empirical evidence for a feedback effect.

An independent checker reads the original CSV using the standard-library parser, reconstructs eligibility with integer comparisons, and calculates rates and scalar influence sums without importing the primary routines or their saved sufficient statistics. It matches every joint cell, row-eligibility digest, cluster count, covariance entry, interval and date deletion across 84 checks. Numerical agreement establishes reproducibility of the implemented calculation, within its declared tolerances. As Peng (2011) distinguishes, an inspectable computation remains different from independent replication of the underlying research setting. Accountable human approval of source claims, authorship, declarations and submission remains pending.

In this selected random-exposure archive, below-three-second records contain less recorded hate and less recorded like than longer records. The directions survive the complete fixed menu of date deletions, and the sparse simultaneous flags remain included. The finding answers a signal-comparison question. Determining whether short viewing reflects preference, playback trouble, an interrupted interaction or another cause requires evidence beyond the three measured fields. Preserving that boundary allows the observed negative hate contrast to remain informative without turning it into an unsupported account of user intent.

References

1. Cameron, A. Colin; Gelbach, Jonah B.; Miller, Douglas L.. (2011). Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics. https://doi.org/10.1198/jbes.2010.07136.
2. Hunter, John D.. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. https://doi.org/10.1109/mcse.2007.55
3. McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. Proceedings of the Python in Science Conference. https://doi.org/10.25080/majora-92bf1922-00a
4. Gebru, Timnit; Morgenstern, Jamie; Vecchione, Briana; Vaughan, Jennifer Wortman; Wallach, Hanna; Daumé III, Hal; Crawford, Kate. (2021). Datasheets for datasets. Communications of the ACM. https://doi.org/10.1145/3458723
5. Rendle, Steffen; Freudenthaler, Christoph; Gantner, Zeno; Schmidt-Thieme, Lars. (2009). BPR: Bayesian Personalized Ranking from Implicit Feedback. Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. PMLR R7 (reissued 2026). https://proceedings.mlr.press/r7/rendle09a.html
6. Hu, Yifan; Koren, Yehuda; Volinsky, Chris. (2008). Collaborative Filtering for Implicit Feedback Datasets. 2008 Eighth IEEE International Conference on Data Mining. https://doi.org/10.1109/icdm.2008.22
7. Zhao, Zhe; Hong, Lichan; Wei, Li; Chen, Jilin; Nath, Aniruddh; Andrews, Shawn; Kumthekar, Aditee; Sathiamoorthy, Maheswaran; Yi, Xinyang; Chi, Ed. (2019). Recommending what video to watch next: a multitask ranking system. Proceedings of the 13th ACM Conference on Recommender Systems. https://doi.org/10.1145/3298689.3346997
8. Herlocker, Jonathan L.; Konstan, Joseph A.; Terveen, Loren G.; Riedl, John T.. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems. https://doi.org/10.1145/963770.963772
9. Pan, Yunzhu; Li, Nian; Gao, Chen; Chang, Jianxin; Niu, Yanan; Song, Yang; Jin, Depeng; Li, Yong. (2023). Learning and Optimization of Implicit Negative Feedback for Industrial Short-video Recommender System. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. https://doi.org/10.1145/3583780.3615482
10. Kassis, Timothy; Agarwal, Vinayak; He, Yuhuan; Patel, Darshil; Brueckner, Aubrey M.. (2026). Scientific Agent Skills: A Library of Procedural Knowledge for Research Agents. arXiv. https://doi.org/10.48550/arXiv.2609.00065

11. Claypool, Mark; Le, Phong; Wased, Makoto; Brown, David. (2001). Implicit interest indicators. Proceedings of the 6th international conference on Intelligent user interfaces. https://doi.org/10.1145/359784.359836

12. KuaiRand project. (n.d.). KuaiRand release README, field definitions, license and scenario discussion. Official repository and issue 3. Accessed 17 November 2026. https://github.com/chongminggao/KuaiRand

13. Yuan, Yuewei; Xu, Tianyi; Huang, Jialei; Yin, Jiayang. (2026). Playback Quality Anomalies Associated with Rapid Swiping in Short Videos: Model-attributed Association Estimation and Risk Prediction. SSRN working paper 7393398. https://doi.org/10.2139/ssrn.7393398.

14. Wang, Richard Y.; Strong, Diane M.. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. https://doi.org/10.1080/07421222.1996.11518099

15. Cameron, A. Colin, and Douglas L. Miller. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.

16. Cena, Federica; Console, Luca; Vernero, Fabiana. (2023). How to deal with negative preferences in recommender systems: a theoretical framework. Journal of Intelligent Information Systems. https://doi.org/10.1007/s10844-022-00705-9

17. Joachims, Thorsten; Granka, Laura; Pan, Bing; Hembrooke, Helene; Gay, Geri. (2005). Accurately interpreting clickthrough data as implicit feedback. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. https://doi.org/10.1145/1076034.1076063

18. Xie, Ruobing; Ling, Cheng; Wang, Yalong; Wang, Rui; Xia, Feng; Lin, Leyu. (2020). Deep Feedback Network for Recommendation. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. https://doi.org/10.24963/ijcai.2020/349

19. Oard, Douglas W.; Kim, Jinmook. (1998). Implicit Feedback for Recommender Systems. AAAI Workshop on Recommender Systems. https://terpconnect.umd.edu/~oard/pdf/aaai98.pdf

20. Jawaheer, Gawesh; Szomszor, Martin; Kostkova, Patty. (2010). Comparison of implicit and explicit feedback from an online music recommendation service. Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems. https://doi.org/10.1145/1869446.1869453

21. Marlin, Benjamin M.; Zemel, Richard S.. (2009). Collaborative prediction and ranking with non-random missing data. Proceedings of the third ACM conference on Recommender systems. https://doi.org/10.1145/1639714.1639717

22. Schnabel, Tobias; Swaminathan, Adith; Singh, Ashudeep; Chandak, Navin; Joachims, Thorsten. (2016). Recommendations as Treatments: Debiasing Learning and Evaluation. Proceedings of the 33rd International Conference on Machine Learning, PMLR 48, 1670-1679. https://proceedings.mlr.press/v48/schnabel16.html

23. Gao, Chongming; Li, Shijun; Zhang, Yuan; Chen, Jiawei; Li, Biao; Lei, Wenqiang; Jiang, Peng; He, Xiangnan. (2022). KuaiRand: An Unbiased Sequential Recommendation Dataset with Randomly Exposed Videos. Proceedings of the 31st ACM International Conference on Information & Knowledge Management. https://doi.org/10.1145/3511808.3557624

24. Peng, Roger D.. (2011). Reproducible Research in Computational Science. Science. https://doi.org/10.1126/science.1213847

25. Campbell, Donald T.; Fiske, Donald W.. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. https://doi.org/10.1037/h0046016

26. Hand, David J.. (2004). Measurement Theory and Practice: The World Through Quantification. Wiley. https://www.wiley-vch.de/de/fachgebiete/mathematik-und-statistik/measurement-theory-and-practice-978-0-470-68567-9

27. Jacobs, Abigail Z.; Wallach, Hanna. (2021). Measurement and Fairness. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3442188.3445901

28. Freedman, David A. (2006). On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors". The American Statistician. https://doi.org/10.1198/000313006x152207

29. Harris, Charles R.; Millman, K. Jarrod; van der Walt, Stéfan J.; Gommers, Ralf; Virtanen, Pauli; Cournapeau, David; Wieser, Eric; Taylor, Julian; Berg, Sebastian; Smith, Nathaniel J.; Kern, Robert; Picus, Matti; Hoyer, Stephan; van Kerkwijk, Marten H.; Brett, Matthew; Haldane, Allan; del Río, Jaime Fernández; Wiebe, Mark; Peterson, Pearu; Gérard-Marchant, Pierre; Sheppard, Kevin; Reddy, Tyler; Weckesser, Warren; Abbasi, Hameer; Gohlke, Christoph; Oliphant, Travis E.. (2020). Array programming with NumPy. Nature. https://doi.org/10.1038/s41586-020-2649-2.

30. McNee, Sean M.; Riedl, John; Konstan, Joseph A.. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. CHI '06 Extended Abstracts on Human Factors in Computing Systems. https://doi.org/10.1145/1125451.1125659