Appendix 1. Appendix

Appendix 1.1 Appendix A: Reliability Tests

For intercoder reliability, each author re-coded 25 randomly selected studies, 9 that were not included and 14 that were included in the study. Regarding intracoder reliability, the authors also coded 25 randomly selected studies, that they did not yet code. Reliability was coded with Krippendorff's alpha, however, as some variables were only present in very few instances in the reliability sample, we also added pairwise agreement (in percentages) to illustrate the agreement, when Krippendorffs' alpha is around or below 0.6. We did not include reliability tests for variables v1 (authors), v2 (year of publication), and v3 (publication venue) due to them not being coded but directly adopted from the citation manager. Concerning the variables on the text corpora, we did not test reliability on v6 (name), v7(originality), and v10(genre) as this information was added after the initial coding procedure, by one of the authors. Which topic modeling method was applied in the study (v11) was coded inductively and thus not included in the reliability analysis. The same is true for the validation method (v12), however, as this is the central variable of this study, we added reliability tests, on the category level to ensure the quality of our results. The lowest agreement score is the inter reliability on the validation category of "comparing methods and hyperparamters". To mend this the authors went over all methods in this category one by one and discussed the coding scheme of each of them. The authors found that the disagreement was limited to one validation method "splitting documents" and thus this validation method was recoded for each of the articles.

Table 1. Overivew of Intercoder and Intracoder Reliability

	Intracoder Author 1	Intracoder Author 2	Intercoder
Exclusion	1	1	1
Substantive RQ	0.87	1	0.65 (84.4%)
Methodological focus	0.87	1	0.51(81.3%)
Error Rate Analysis	0.87	1	0.87 (75%)
Qualitative Interpretation (Internal and External)	0.48 (86.6%)	0.00 (93.3%)	0.69
Downstream Tasks	0.77	0.85	0.39 (75%)
Comparing Models	0.87	0.48 (73.3%)	0.09 (62.5%)
Information Theory Metrics	0.64 (93.3%)	1	0.92
Similarity and Distance Measures	0.77	0.64 (93.3%)	0.00 (78.1%)

Appendix 1.2 Appendix B: Overview of Validation Methods

Table 2. Classification of all Validation Methods included in this Study, with the total number of application and studies, in which it is applied

Validation Method	application	studies
Model Comparison	**763**	**483**
Cross-Validation	97	94
Applying different Methods	214	214
Split Train Test Set	305	303
Baseline Model	147	128
Distinctivness of Topwords	**274**	**176**
Coherence Scores	234	169
Exclusivity	26	26
Purity	14	14
Downstream Tasks	**334**	**320**
Error Rate Analysis	**618**	**351**
Accuracy	139	137
Area Under the ROC Curve (AUC-ROC)	30	25
Error 1 and 2	22	12
F-Score	139	135
Mean Absolute Error (MAE)	11	10
Mean Squared Error (MSE)	9	9
Precision and Recall	243	199
Recall-Oriented Understudy for Gisting Evaluation (ROUGE)	11	11
Root Mean Squared Error (RMSE)	14	13
Internal Qualitative Inspection	**634**	**428**
Consulting Topic Experts for Evaluation	36	35
Topic Interpretation	211	187
Reading Top Documents	51	51
Topic Labeling	320	310
Word Intrusion	16	16
External Qualitative Inspection	**210**	**177**
Comparison with inductive corpus coding	85	77
Theoretical Considerations	76	75
Real Life Dynamics / external events	49	48
Information Theory Metrics	**223**	**191**
Entropy	11	11
Jensen-Shannon Divergence (JSD)	26	24
Kullback–Leibler Divergence (KL)	37	36
Perplexity	149	144
Similarity and Distance Metrics	**80**	**66**
Jaccard Coefficient	13	13
Silhouette	15	14
Similarity	52	45

Appendix 1.3 Appendix C: Further Information

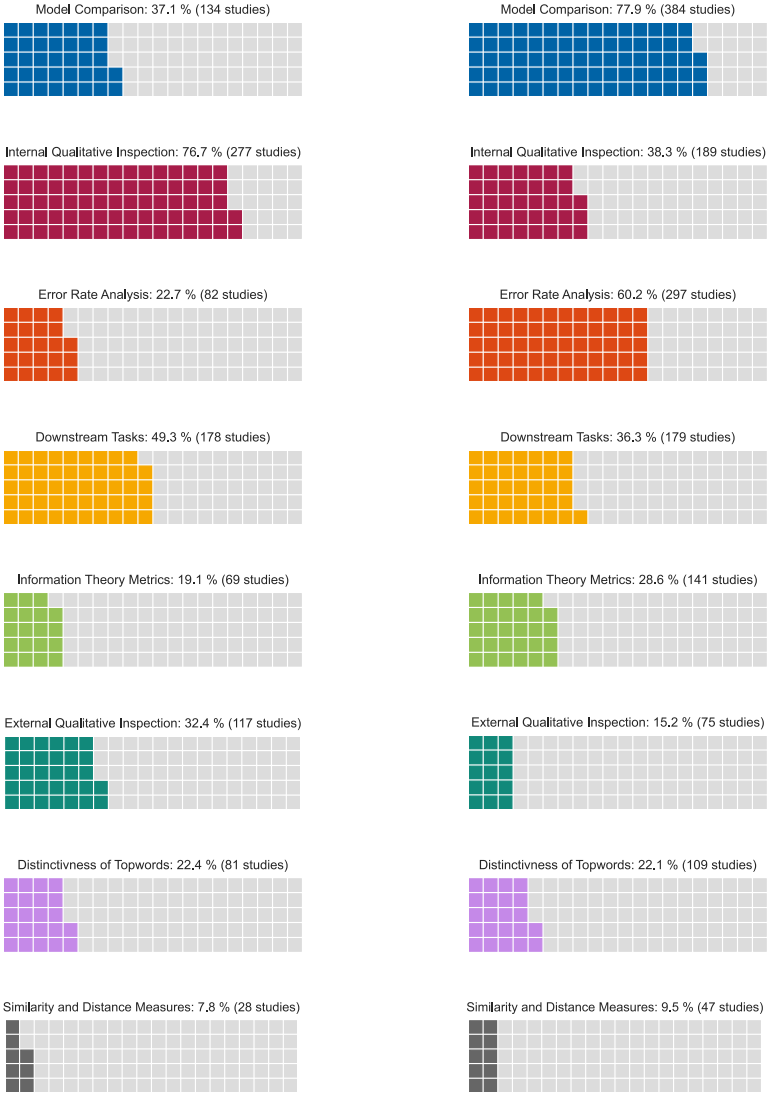


Figure 7. Percentage of substantive (left panel) and methodological (right panel) Studies employing validation methods

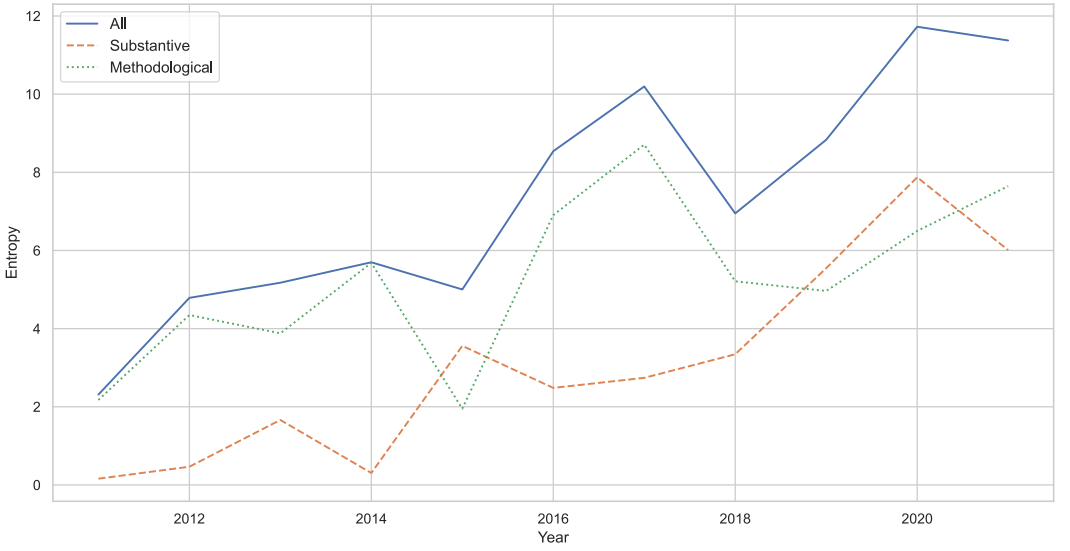


Figure 8. Information Entropy for validation methods over time

Top Publication Outlets	
ACM International Conference on Information and Knowledge Management	38
International Conference on World Wide Web	34
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	25
ACM SIGIR Conference on Research and Development in Information Retrieval	22
ACM International Conference on Web Search and Data Mining	18
ACM Transactions on Knowledge Discovery from Data	9
Conference on Empirical Methods in Natural Language Processing	9
ACM Transactions on Intelligent Systems and Technology	9
IEEE ACCESS	9
Journal of Machine Learning Research	8

Top Publication Journals	
ACM Transactions on Intelligent Systems and Technology	9
IEEE ACCESS	9
ACM Transactions on Knowledge Discovery from Data	9
Journal of Machine Learning Research	8
International Journal of Communication	8
ACM Transactions on Information Systems	8
IEEE/ACM Transactions on Audio, Speech, and Language Processing	7
Communication Methods & Measures	7
Political Communication	6
Marketing Science	6

Top Publication Conferences	
ACM International Conference on Information and Knowledge Management	38
International Conference on World Wide Web	34
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	25
ACM SIGIR Conference on Research and Development in Information Retrieval	22
ACM International Conference on Web Search and Data Mining	18
Conference on Empirical Methods in Natural Language Processing	9
ACM Conference on Web Science	7
ACM Conference on Computer Supported Cooperative Work and Social Computing	5
IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining	5
Annual Meeting of the Association for Computational Linguistics	5

Top Publication Outlets Core Social Science	
International Journal of Communication	8
Communication Methods & Measures	7
Political Communication	6
Marketing Science	6
Environmental Communication	4
Journalism Studies	4
Journal of Broadcasting & Electronic Media	3
International Conference on Social Media and Society	2
American Sociological Review	2
International Conference on Digital Government Research	2

Top Publication Outlets Peripheral Social Science	
ACM International Conference on Information and Knowledge Management	38
International Conference on World Wide Web	34
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	25
ACM SIGIR Conference on Research and Development in Information Retrieval	22
ACM International Conference on Web Search and Data Mining	18
ACM Transactions on Knowledge Discovery from Data	9
ACM Transactions on Intelligent Systems and Technology	9
IEEE ACCESS	9
Conference on Empirical Methods in Natural Language Processing	9
ACM Transactions on Information Systems	8

Table 3. Overview of Top Publication Outlets of Studies in our Systematic Literature Review

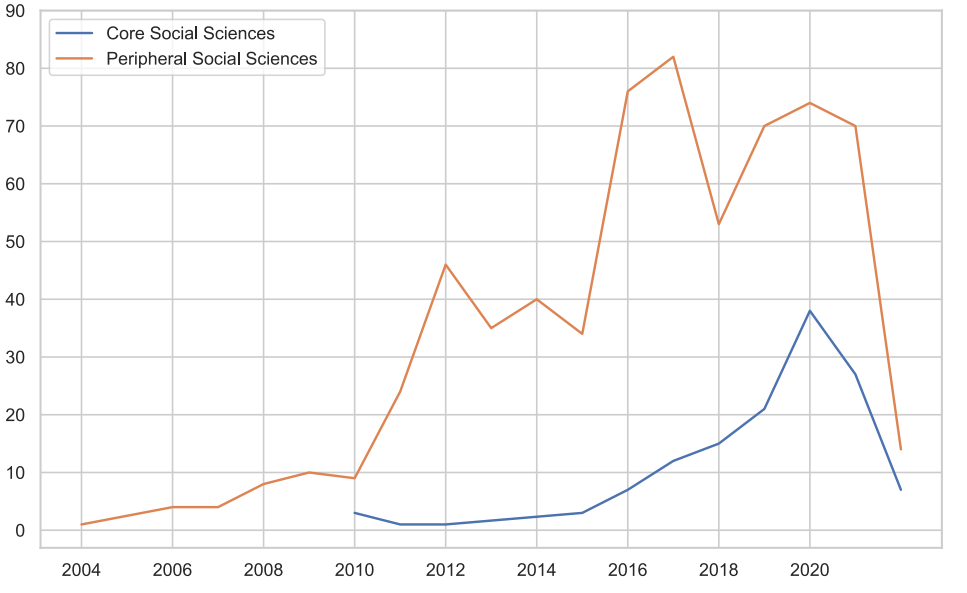
Appendix 1.4 Appendix D: Robustness Analysis

In our literature review on topic modeling validation, we sought to enhance the robustness of our findings by conducting a pooled analysis that differentiates between social science and non-social science studies. Recognizing that methodological and thematic variances might influence the outcomes across diverse research domains, we categorized the publication outlets into these two broad categories. By re-running our entire analysis strategy separately for social science and non-social science studies, we aimed to investigate whether our original conclusions held consistent across different academic disciplines.

This additional analysis serves as a robustness check, ensuring that our findings are not biased by the nature of the publication outlets. The results, presented in the appendix, demonstrate a lack of convergence in topic modeling validation practices across both social science and non-social science studies. This reinforces the validity of our original findings and underscores the widespread challenges in achieving consistent and reliable validation in topic modeling, regardless of the disciplinary context.

Additionally, we conducted a second robustness check by applying a weighted analysis based on citation counts adjusted for publication year. This approach accounted for differences in citation accumulation across older and more recent studies, offering a way to evaluate if the number of citations influenced the trends in validation practices. The results of this weighted analysis, also included in the appendix, showed that while our overall conclusions still hold, there were some differences, particularly a slight decrease in entropy for social science studies in the final year of our dataset. However, given the small sample of 27 studies in this case, we advise caution in interpreting this result as a broader trend.

Together, these two robustness checks confirm the stability of our findings, strengthening our original conclusions about the lack of convergence in topic modeling validation practices.



Note: As only a quarter of the year 2022 is included in the sample we did not include it in the graph, as it would have resulted in a misleading trend.

Figure 9. Number of Studies from the Core Social Sciences vs. Peripheral Social Science in our sample over time.

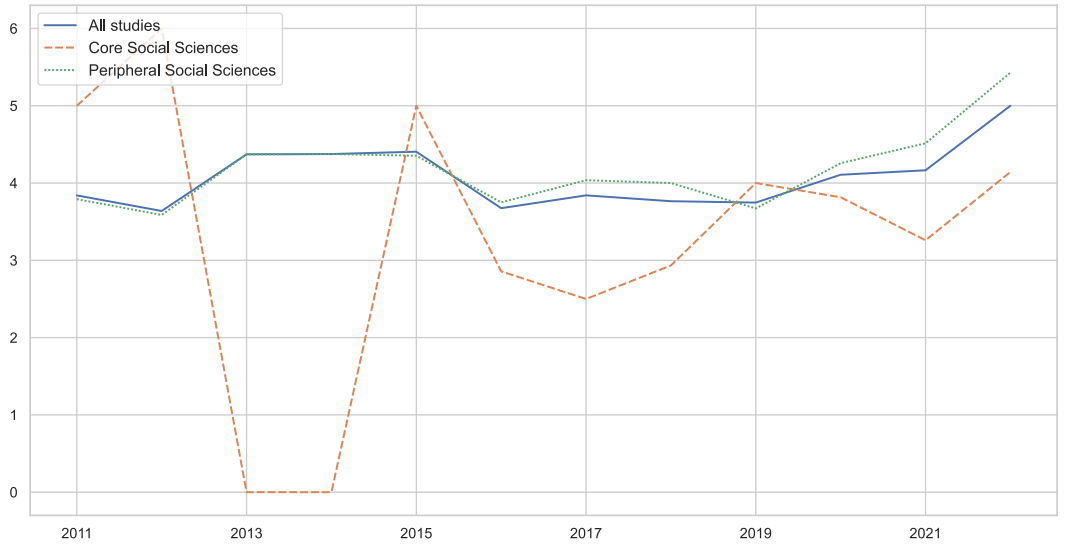


Figure 10. Average number of validation categories used per study over time Core Social Sciences vs. Peripheral Social Science.

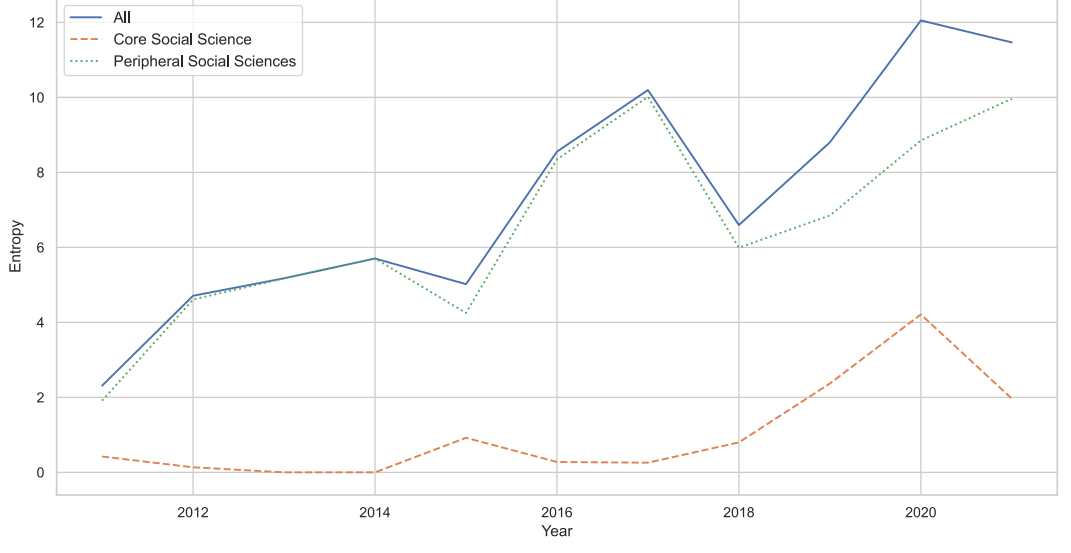


Figure 11. Information Entropy for validation methods over time Core Social Sciences vs. Peripheral Social Science

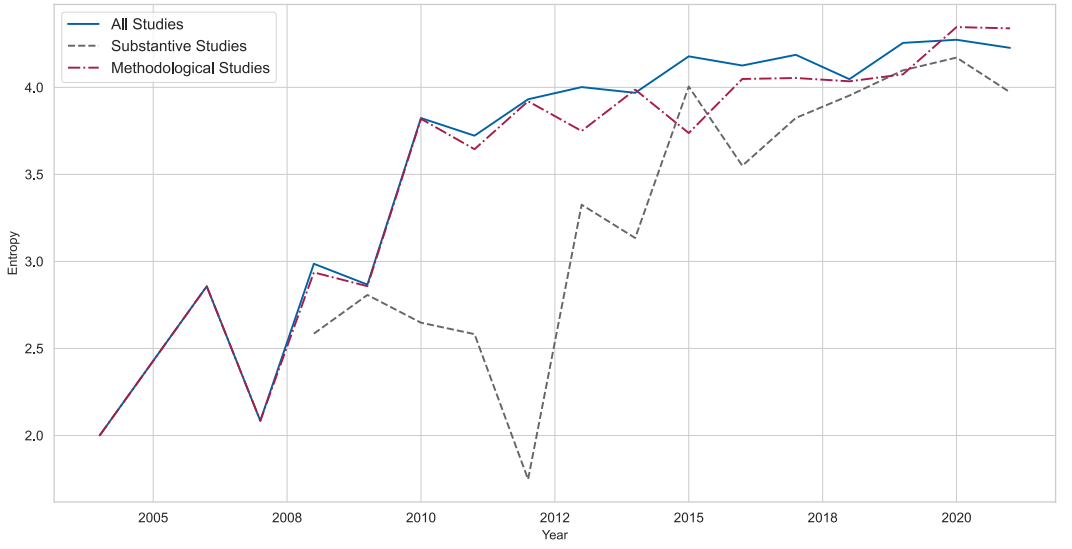


Figure 12. Weighted Information Entropy for validation methods over time Core Social Sciences vs. Peripheral Social Science

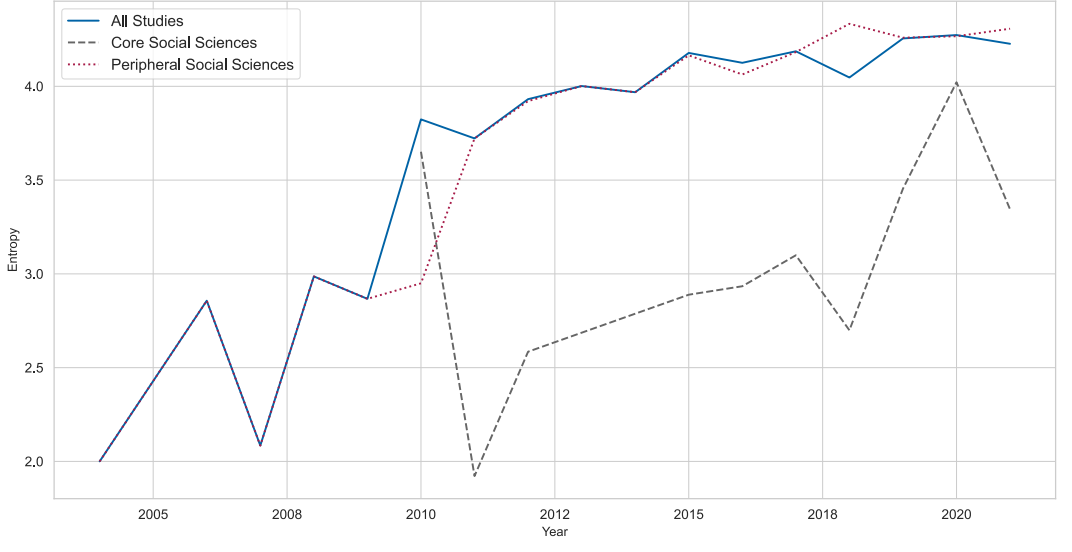


Figure 13. Weighted Information Entropy for validation methods over time Core Social Sciences vs. Peripheral Social Science