Dissertation Paper I | Draft

Working title: **Mapping and Evaluating Measurement Validity Approaches for Text-as-Data Methods**

Lukas Birkenmaier  
MA Politik- und Verwaltungswissenschaften (University of Konstanz) &   
MSc Big Data and Digital Futures (University of Warwick)

***Coauthors:***

1. Dr. Clemens Lechner

2. Prof. Dr. Claudia Wagner

# Background

## The rise of computational text analysis methods

In the past decade, social science research has experienced a remarkable rise in publications relying on digital behavioral data and computational methods (Brady, 2019; Edelmann et al., 2020). Thereby, the analysis of texts using computational text analysis methods (CATM) holds particular promise, as textual data enable researchers to facilitate substantively important inferences of human behavior on an unprecedented scale (Grimmer & Stewart, 2013; Lazer et al., 2009).

* Vorteile und potentiale Textdaten
* Herausforderungen, ethische bedenken bei falscher messung

## The critical tasks of method validation text as data research

In reaction to the popularity of automated analysis of textual data, a growing body of research highlights several methodological challenges in the automated analysis of textual data (Baden et al., 2021; Howison et al., 2011; van Atteveldt et al., 2021). One of the key challenges relates to the validity of CATM-based measures. Several scholars have expressed doubts about whether CATM can operationalize and validly measure the constructs under study, thus questioning these methods’ ability to answer substantive theory-driven research questions (Adcock & Collier, 2001). This is especially relevant because the high dimensionality and complex structure of texts renders it almost impossible to *truly* understand the underlying dependencies of texts (Yeomans, 2021).[[1]](#footnote-2)[[2]](#footnote-3) This culminates in the problem that even hand-coded estimates (“gold standard”) for the same texts are often [not consistent](https://www.linguee.de/englisch-deutsch/uebersetzung/not+consistent.html) so that reliability measures can vary significantly between different coders (cf. Song et al., 2020). Therefore, authors such as Baden et al. (2021) have called for more unifying validation efforts to show “exactly how and how convincingly [CATM] operationalize relevant conceptual properties” (Baden et al., 2021, p. 14). When reviewing the literature, one can observe a great heterogeneity of validation approaches in studies using CATM developed over time. Among others, these approaches include the comparison of estimates with other measurements and external data (Bach et al., 2021; Kmetty & Németh, 2022; Röttger et al., 2021), or the development of multistage validation frameworks, which cover different dimensions of validity at the same time (Goet, 2019; Lowe & Benoit, 2013). Even more worrying, scholars who want to validate their measures can easily lose track of what options they have, and which validation approaches are necessary to conduct high-quality research projects.

* Thus, it is asked: How are issues of measurement validity addressed in CATM studies? What are the main validation strategies in CATM research? And what are possible solutions heterogeneity of approaches

# Research Design

To assess validation practices in text as data research, we pursue a two-fold approach. First, a systematic review of validation practices in peer-review publications. And second, expert interviews with outstanding scholars in the field of text analysis for political communication research.

## Systematic Review

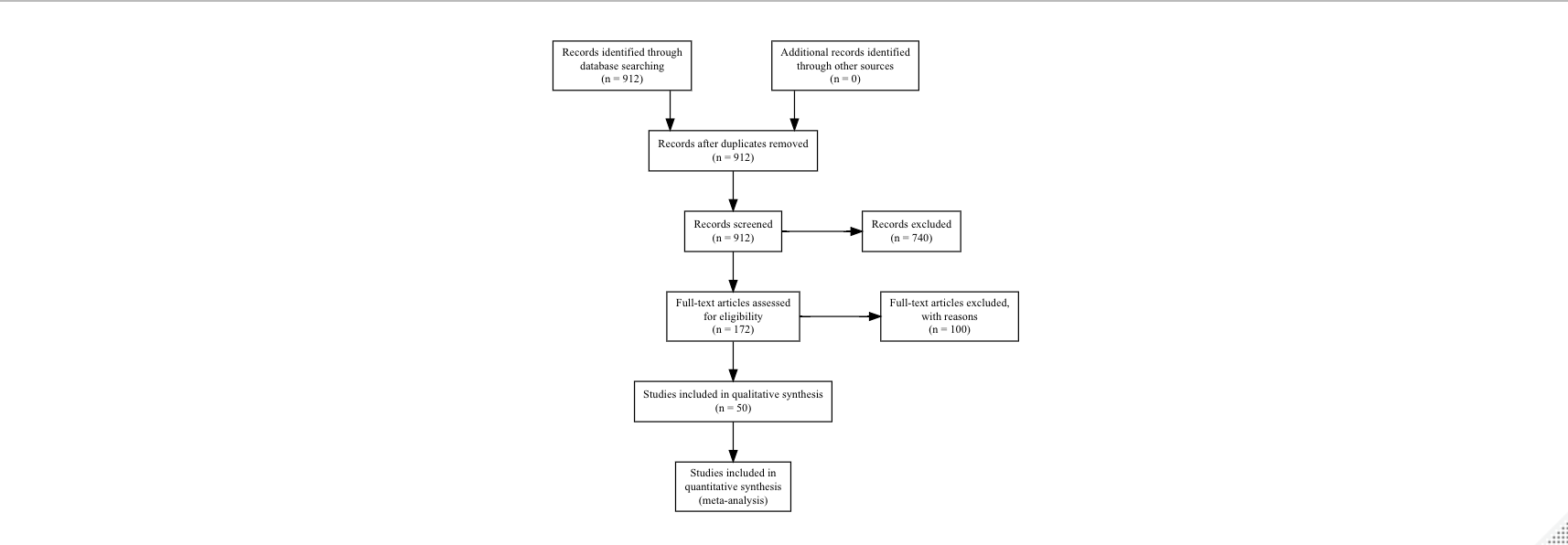
To conduct the review, we use a systematic strategy to transparently identify relevant publications for CATM research (Durlak & Lipsey, 1991; Liberati et al., 2009). To do so, we rely on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA), a systematic search and evaluation procedure of academic literature (Moher et. al., 2009). Figure 2 provides an overview of the review process.

Figure 1: PRISMA Statement

### Literature Search

We decide to limit our literature search to publications in the field of political communication. Generally, the field of political communication has experienced remarkable methodological advancements in the last decade, with scholars deploying sophisticated analytical methods and driving the development of CATM (Theocharis & Jungherr, 2021, p. 2). In addition, the field of political communication promises to gain further relevance in the future, because it deals with some of the most pressing challenges facing societies today, such as digital dynamics, fake news, and political polarization (Dufva & Dufva, 2019).   
To limit our potential scope of publications, we identify the top five most cited peer-reviewed journals in the field of “political science research” and “communication research”[[3]](#footnote-4). Given the innovative character of many CATM publications, however, we further identified six journals with a proven record of publishing high-quality text as data studies. Thus, this leaves us with a list of 16 relevant journals (see Appendix 1).

To search for relevant publications within these journals, we first started a naïve search strategy using terms related to the data type (“text\*”) and field of research (“poli\* OR part\* OR govern\*). After a first search round, we furthermore optimize our search strategy using computational methods for automated search term selection (QUELLE LITSEARCHR). The final search string can be found in Appendix 2. Because we want to describe current validation practices in CATM research, we only include studies published after 2014. Ultimately, this results in a preliminary literature list of 920 publications.

Because we are primarily interested in studies which apply a CATM to measure (unobservable) constructs based on textual data, a second screening round was conducted by examining the title, abstract and keywords of all studies identified.[[4]](#footnote-5) Thereby, inclusion and exclusion criteria were the following:

* Publication must apply a **computer-assisted text method** (e.g., social media entries, newspaper articles, speech transcripts, parliament documents, etc.)
* Publication must be focused on a research topic in the field of **political communication**
* Publication must conduct substantial and theory-driven **empirical work** (not only method discussion, development etc.)

Ultimately, this reduced the preliminary literature list to 172 publications which meet the eligibility criteria and were included in the review.

### Coding Procedure

Each of the articles is then coded using a coder manual. In a first step, we collect general information for the studies, such as data type and language(s), method type and construct of interest. In a second step, we code the validation steps reported in each publication. A validation step is defined as a complete and self-contained validation exercise, such as e.g., comparing the output of a CATM with hand-coded scores, or using the CATM scores to predict an external but related criterion. If there is an explicit link to the Appendix, we also inspected and code supplementary materials. For each validation step, we copy the whole section in the paper relating to the validation exercise into our list and record how the authors themselves referred to the respective validation exercise. In addition, we also assign each validation exercise a validation type. Keeping the absence of a general framework to distinguish and classify different types of validation exercise in mind (Grimmer et al., 2022), we initially differentiate between two major groups of approaches. First, the direct comparison and cross-examination of CATM estimates with other data (either hand-coded scores, scores from other CATM, or external criterions). And second, validation exercises related to the human evaluation on the plausibility of the method and its measures. Furthermore, we also provide an additional option when a validation step is not clearly assignable. In a last step, we take a more inductive approach and aim to subdivide and structure the validation approaches into more granular categories. This is done by a discussion between the authors, having the goal to map applied validation strategies and grouping them into distinguishable groups of approaches.

To ensure the quality of our coding manual, two coders coded a subset of 20% of the studies (see Landis & Koch, (1977)). The coding Interrater Reliability reached sufficient congruence (XX).[[5]](#footnote-6) initial search results (a sufficient sample size according to Riffe, Lacy, and Fico (2014, 111), and indicated very good agreement (Cohen’s j ¼.83; Landis and Koch 1977).

Intercoder reliability calculations regarding the reliability of the coding procedure between the first author and a student assistant with expertise in the field research based on a random sample of 100 publications from the

* Exclusion criteria: often only description of text attributes, such as word length or activity under a post (e.g. (Koc-Michalska, n.d.))

## Expert Interviews

. To identify relevant interview partners, we

* “ Experts have more than just systematic organized knowledge, they also have deep knowledge in specific experiences which result from their actions, responsibilities, obligations of the specific functional status within an organization. Similarly, [Meuser and Nagel (1991)](https://www.sciencedirect.com/science/article/pii/S0740624X18304131" \l "bb0205) see an expert as a person responsible for a concept, an implementation or ability to solve a problem, as someone who has relevant factual knowledge, aggregated or specific knowledge about processes, group behaviors, strategic decisions but also has knowledge, (general) information or privileged access to information. This type of knowledge is often implicit or difficult to articulate, and therefore we need a specific approach to interviewing in order to access the experts' knowledge. With the use of expert interviews, we are not interested in their individual biographies, but in their viewpoints, and, as representatives of a larger domain, such as the organization, to their privileged access to decision-making processes and people.”

### Selection

### Interviews

The expert interviews were conducted either face-to-face or using online video tools (Microsoft Teams).

* Vor- und Nachteile methoden
* Ablauf Interview
  + *“Prior to the interviews, the experts were asked for permission to record the interview for accuracy purposes. Interviewees were ensured anonymity by declaring that their personal data will not be made available to any third parties or made public. At the end of the interview, experts were also asked if they would like to add any comments and insights they thought appropriate to the interview and topic, as well as suggest further experts and potential interviewees.”*

The interviews were designed to investigate the research phenomenon in an open-minded way and to understand how scholars in the field approach issues of CATM validation.

To do so, we choose semi-structured expert interview was chosen for the data collection. According to Helfferich (2022), this is recommended for expert interviews, as it allows questions to be focused more narrowly and makes it easier to structure the interview.

All questions in the guideline were framed in an open-ended nature and no “right” or “wrong” answers were expected. To do so, experts were first asked to provide their own definition of measurement validity in text as data research. Then, they were invited to share their opinion on validation practices in the field.

Furthermore, care was taken during data collection to ensure the privacy of the experts. Thus, individual quotes are only presented anonymously, and central statements of the stakeholders are only reproduced in aggregated form.

### Analysis

To analyze the interviews, the qualitative content analysis method according to Mayring was used (Mayring, 2010). This is particularly useful for structuring texts with the aim of identifying cross-references and similar explanatory patterns (Mayring, 2010, p. 602). The central methodology here is to classify individual statements into categories with the help of codes (Mayring, 2010, p. 603). The categories summarize the experts' statements and help to structure the data (Kuckartz, 2012).

The software program MAXQDA 11 for computer-assisted qualitative data and text analysis was used to code the interviews

# Results

## Systematic Review

* Validation across groups of methods
  + Topic Model and other change?

# Conclusion

* Validity nicht nur Textdaten, sonder alle Formen von DVD

# Analysis & Results

# Conclusion

Notes:

* Often, unclear when face validity begins
  + Considerations about appropriateness of method choice etc. are not systematically reported -> when does face validity begin?

# Appendix

## Appendix 1: Keywords

* + Naïve search
    - (Politi\* OR Party OR Govern\*) AND text\*
  + Adjusted search (*LitsearchR*)
    - Topic
      * **Political Communication** 
        + (elect\* OR govern\* OR parti\* OR polici\* OR "social\* media\*" OR polit\*)
      * **Textual Analysis**
        + ("autom\* text\* analysi\*" OR "content\* analysi\*" OR sentiment\* OR discours\* OR languag\* OR "machin\* learn\*" OR text\* OR word\* OR "comput\* communic\* scienc\*" OR Corpus\* OR lexicon\* OR Automa\* Content\* Analy\*)

## Appendix 2: Journals

* + - Communication
      * Communication Methods and Measures
      * Digital Journalism
      * Political Communication
      * Applied Linguistics
      * Research on Language and Social Interaction
      * Journal of Communication
      * Public Opinion Quartely
      * [Administrative Science Quarterly](https://www.scimagojr.com/journalsearch.php?q=16036&tip=sid&clean=0)
      * [American Sociological Review](https://www.scimagojr.com/journalsearch.php?q=16929&tip=sid&clean=0)
      * [American Political Science Review](https://www.scimagojr.com/journalsearch.php?q=15557&tip=sid&clean=0)
      * [Annual Review of Political Science](https://www.scimagojr.com/journalsearch.php?q=15646&tip=sid&clean=0)
      * [Annual Review of Sociology](https://www.scimagojr.com/journalsearch.php?q=16976&tip=sid&clean=0)
      * [American Journal of Political Science](https://www.scimagojr.com/journalsearch.php?q=15555&tip=sid&clean=0)
      * [Leadership Quarterly](https://www.scimagojr.com/journalsearch.php?q=21149&tip=sid&clean=0)
      * [Political Analysis](https://www.scimagojr.com/journalsearch.php?q=145558&tip=sid&clean=0)

## Appendix 3: Benchmark Literature

* Goet, Niels D. ‘Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015’. *Political Analysis* 27, no. 4 (2019): 518–39.
* Gilardi, Fabrizio, Theresa Gessler, Mael Kubli, and Stefan Muller. ‘Social Media and Political Agenda Setting’. Political Communication 39, no. 1 (2 January 2022): 39–60.
* Di Cocco, J., & Monechi, B. (2022). How Populist are Parties? Measuring Degrees of Populism in Party Manifestos Using Supervised Machine Learning. Political Analysis, 30(3), 311-327. doi:10.1017/pan.2021.29
* Røed, Maiken. "When do political parties listen to interest groups?." Party Politics (2022): 13540688211062832.
* Welbers, K., van Atteveldt, W., Bajjalieh, J., Shalmon, D., Joshi, P. V., Althaus, S., Chan, C.-., Wessler, H. & Jungblut, M. (2022). [Linking event archives to news: A computational method for analyzing the gatekeeping process](https://madoc.bib.uni-mannheim.de/61633/1/Linking%20event%20archives%20to%20news%20a%20computational%20method%20for%20analyzing%20the%20gatekeeping%20process.pdf). *Communication Methods and Measures, 16*(1), 59–78. https://doi.org/10.1080/19312458.2021.1953455

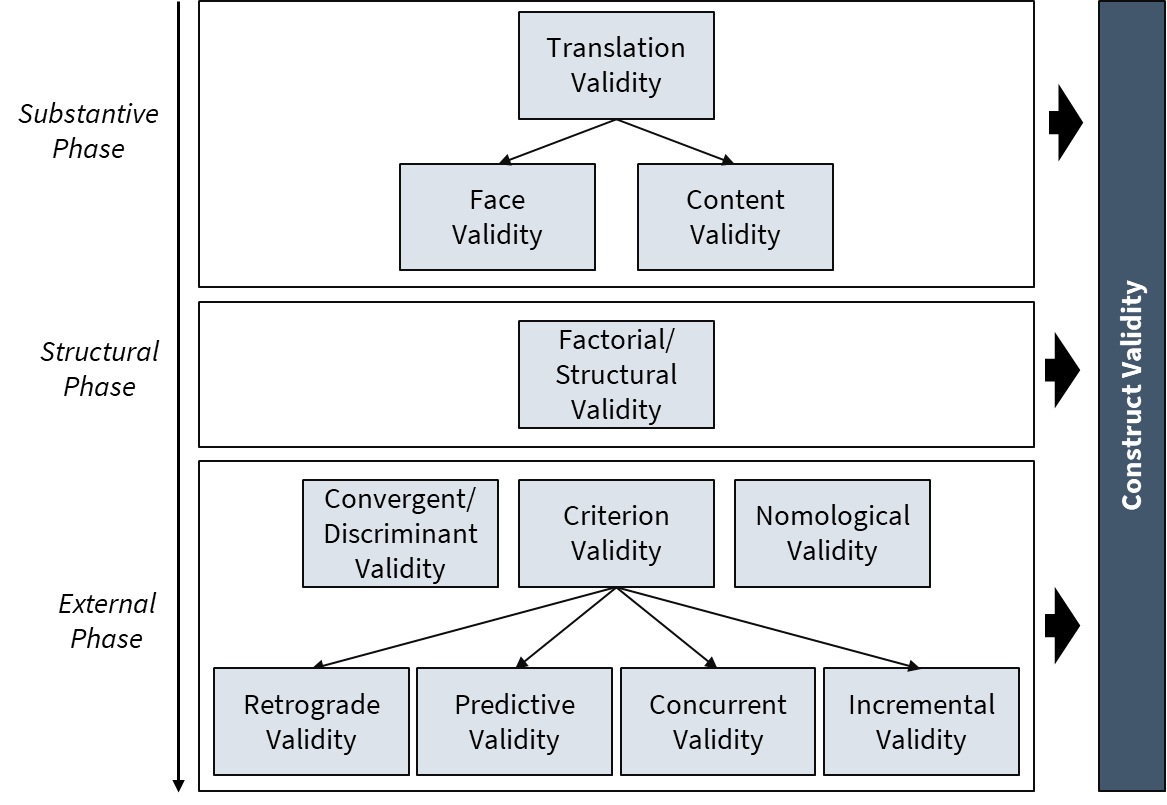
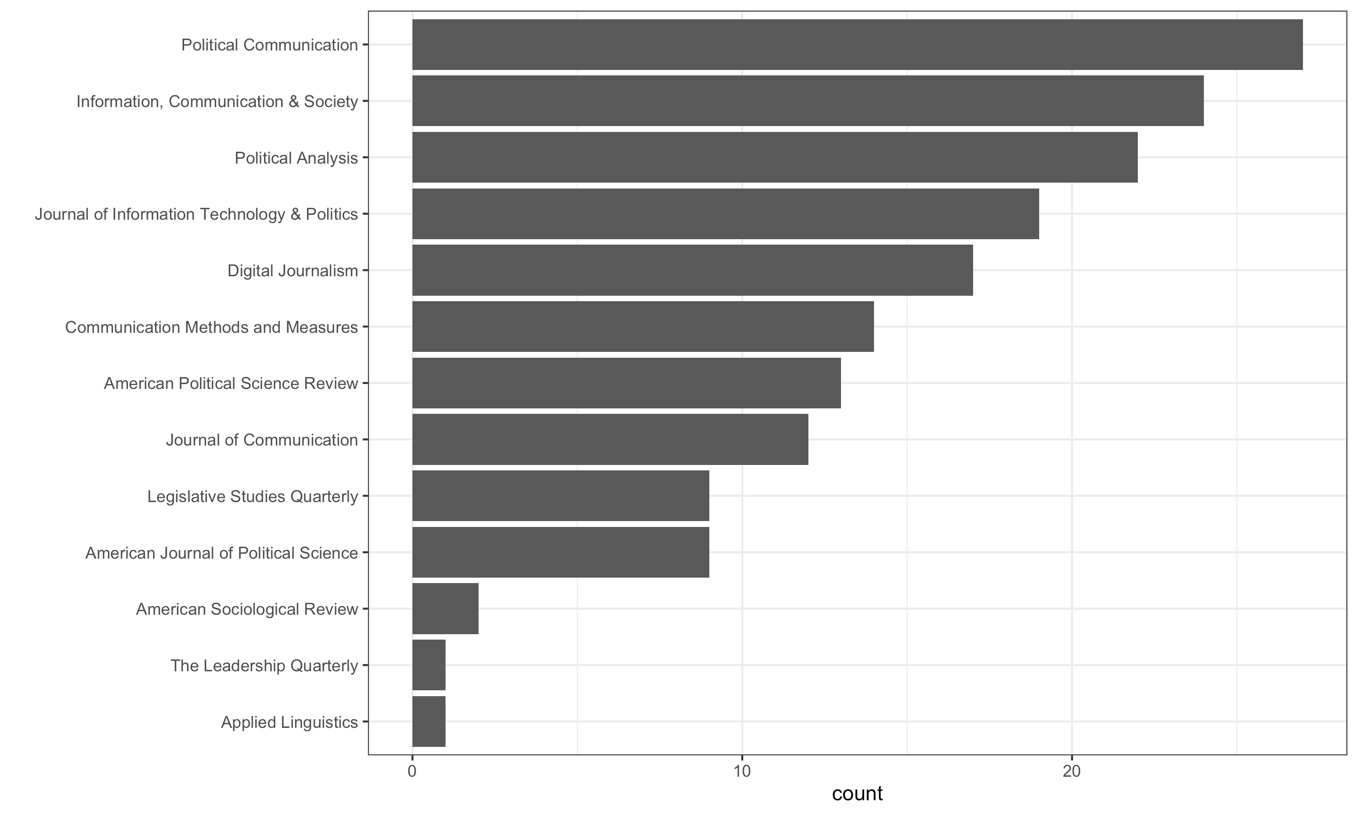
Appendix 3: Overview Validity Concepts

Figure 2: Own visual depiction of validity concepts



## Appendix 4: Questionnaire

|  |  |
| --- | --- |
| Background Paper |  |
|  |  |
| Consent | Furthermore, I would like to inform you that I will record the prepared questions and your answers for my bachelor thesis. This recording will be transcribed to be able to analyze the answers. The recorded data is purely for research purposes Your data will only be presented in aggregated form and anonymized. The research paper, including the appendices, can only be viewed by members of the Department of Geography at Heidelberg University, including myself.  Do you agree that I record this interview for research purposes? |

|  |  |
| --- | --- |
| Introduction | What is your professional background? |
| How is your work related to text analysis and political communication? |
| Measurement Validity | How would you **describe validity**? |
| Have you come across **different validity terms** within different **research fields**? |
| Measurement validity and computer-assisted text methods? | Have you come across **different validity terms** in the field of computer-assisted text methods? |
| How would you describe the process of measurement validation? Which steps are usually reported in the final paper? |
| **What approaches** do you know to validate a computer-assisted text method? |
| How would you **describe** the **current state of knowledge** on validating computer-assisted text method? |
| Challenges Measurement Validity and CATM | What are the barriers that **hamper** the validation of computer-assisted text methods? (*Costs, missing guidelines etc*.) |
| Do you have any ideas on **how to improve** the validation of computer-assisted text methods? |
| Conclusion | Is there anything else you would like to tell us? |

# TOPIC 0 | Summary

**Enhancing the Measurement Validity of Digital Behavioural Data**[[6]](#footnote-7)

**Introduction.** In the past 20 years, social science research has experienced a remarkable rise in publications relying on digital behavioral data and computational methods (Brady, 2019; Edelmann et al., 2020). The potential advantages of these developments are impressive. They allow researchers not only to get access to new kinds of previously unexplored data, but also to study what people think, feel, and do, in near real-time and large scale, by analyzing and combining digital traces of human behaviour (Azucar et al., 2018; Eichstaedt et al., 2021). However, a growing body of research also highlights profound challenges in the validation of DBD-based methods (Baden et al., 2021; Howison et al., 2011; van Atteveldt et al., 2021). One of the most crucial questions is whether methods relying on DBD can operationalize and validly measure the underlying constructs of interest in social science research. Keeping in mind that focal constructs in social-science research are often unobservable and multidimensional, operationalizations that lack validity could seriously hamper these methods’ ability to answer substantive theory-driven research questions because what exactly is being measured is unknown (Flake et al., 2017). This is especially relevant in the context of textual data because the high dimensionality and complex structure of texts render it almost impossible to *truly* understand the underlying dependencies in written language (Song et al., 2020; Yeomans, 2021).

**Dissertation Outline.** Emphasizing the challenges associated with validation practices in DBD research, this dissertation will describe, structure, and advance current validation practices in the field of DBD research. In the first paper, I will conduct a systematic review of validation practices in social-scientific studies that rely on computer assisted text methods (CATM). The aim is to provide an overview over how current studies approach the issue of validity and that, if any, evidence for the validity of their measures they present. , Toward that end, I will identify and structure the most relevant validation practices for a sample of studies that used dictionaries in the domain of political science research. This study will set the groundwork for approaching the issue of measurement validation.

In the second paper, I will build upon the validation practices identified in this systematic review and develop a validation framework, which will synthesize and combine current approaches in the field of DBD research. Additionally, the framework will benefit from a discussion of best practice validation approaches in other domains, such as psychometrics Adcock & Collier, 2001). The third and fourth paper will then conduct empirical research in the field of DBD that showcase measurement validation practices while also addressing substantive questions. In the third paper, I will explore determinants of political polarization and partisan language using textual data. Even though the exact research question is still open for discussion, the paper will include a thorough validation of the measurement using the validation framework developed in paper 1 and 2. In the fourth paper, I want to extent the scope of analysis by assessing the validity of methods using other forms of DBD. Among others, this might include sensor, image, or video data to answer substantive research questions.

**Expected findings.** Results from my dissertation will not only contribute to a better understanding of current validation practices for DBD methods but also help advance the standars of validity and validation practices in the field. Starting from a discussion of validation practices and their associated methodological and theoretical challenges (*descriptive part*), the dissertation will then provide a general validation framework and hands-on guidance on how to convincingly validate DBD based measures in applied research projects (*normative part)*.

Literature

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, *95*(3), 529–546.

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/J.PAID.2017.12.018

Bach, R. L., Kern, C., Amaya, A., Keusch, F., Kreuter, F., Hecht, J., & Heinemann, J. (2021). Predicting voting behavior using digital trace data. *Social Science Computer Review*, *39*(5), 862–883.

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2021). Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. *Communication Methods and Measures*, 1–18. https://doi.org/10.1080/19312458.2021.2015574

Brady, H. E. (2019). The Challenge of Big Data and Data Science. *Annual Review of Political Science*, *22*(1), 297–323. https://doi.org/10.1146/annurev-polisci-090216-023229

Dufva, T., & Dufva, M. (2019). Grasping the future of the digital society. *Futures*, *107*, 17–28. https://doi.org/10.1016/j.futures.2018.11.001

Durlak, J. A., & Lipsey, M. W. (1991). A practitioner’s guide to meta-analysis. *American Journal of Community Psychology*, *19*(3), 291–332.

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, *46*(1), 61–81. https://doi.org/10.1146/annurev-soc-121919-054621

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., & Buffone, A. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, *26*(4), 398.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Goet, N. D. (2019). Measuring polarization with text analysis: Evidence from the UK House of Commons, 1811–2015. *Political Analysis*, *27*(4), 518–539.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Helfferich, C. (2022). Leitfaden-und experteninterviews. In *Handbuch Methoden der empirischen Sozialforschung* (pp. 875–892). Springer.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, *12*(12), 2.

Kmetty, Z., & Németh, R. (2022). Which is your favorite music genre? A validity comparison of Facebook data and survey data. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 07591063211061754. https://doi.org/10.1177/07591063211061754

Koc-Michalska, J., K; Lilleker, DG; Michalski, T; Gibson, R; Zajac. (n.d.). Facebook affordances and citizen engagement during elections: European political parties and their benefit from online strategies? *JOURNAL OF INFORMATION TECHNOLOGY & POLITICS*. http://dx.doi.org/10.1080/19331681.2020.1837707

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., & Gutmann, M. (2009). Computational social science. *Science*, *323*(5915), 721–723.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lowe, W., & Benoit, K. (2013). Validating estimates of latent traits from textual data using human judgment as a benchmark. *Political Analysis*, *21*(3), 298–313.

Röttger, P., Vidgen, B., Nguyen, D., Waseem, Z., Margetts, H., & Pierrehumbert, J. (2021). HateCheck: Functional Tests for Hate Speech Detection Models. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 41–58. https://doi.org/10.18653/v1/2021.acl-long.4

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. G. (2020). In Validations We Trust? The Impact of Imperfect Human Annotations as a Gold Standard on the Quality of Validation of Automated Content Analysis. *Political Communication*, *37*(4), 550–572. https://doi.org/10.1080/10584609.2020.1723752

Theocharis, Y., & Jungherr, A. (2021). Computational Social Science and the Study of Political Communication. *Political Communication*, *38*(1–2), 1–22. https://doi.org/10.1080/10584609.2020.1833121

van Atteveldt, W., van der Velden, M. A., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Communication Methods and Measures*, *15*(2), 121–140.

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012). Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 819–824.

Yeomans, M. (2021). A concrete example of construct construction in natural language. *Organizational Behavior and Human Decision Processes*, *162*, 81–94. https://doi.org/10.1016/j.obhdp.2020.10.008

1. This has been described as the first principle of textual data analysis by Grimmer and Stewart (2013), namely “all quantitative models of language are wrong -but some are useful). Thus, “the complexity of language implies that all methods necessarily fail to provide an accurate account of the data-generating process used to produce texts” (p. 270) [↑](#footnote-ref-2)
2. “The quantities we seek to estimate from text [...] are fundamentally unobservable” (Lowe & Benoit, 2013, p. 299), [↑](#footnote-ref-3)
3. To identify most cited journals, we used the Scimagojr Journal Ranking (https://www.scimagojr.com/). [↑](#footnote-ref-4)
4. To code the abstracts, we use the a*bstrackr* web interface (Wallace et al., 2012) [↑](#footnote-ref-5)
5. To calculate IRR, we [↑](#footnote-ref-6)
6. Note: In a future version of the expose, I am planning to link the title to a specific area of (political) social science research, such as party contestation or polarisation in political discourse. However, I have not yet decided on a specific subfield yet. [↑](#footnote-ref-7)