

Konzept Version 1 der AG 5 (Inter-Codierer-Reliabilitätstest)

Giulia D'Amico, Regina Eicher, Almira Medaric, Alexandra Muster

24. September 2020

Inhaltsverzeichnis

1	Was? Thema und Fragestellung	2
2	Warum? Stand der Forschung und Relevanz des Inter-Codierer-Reliabilitätstest	2
2.1	Stand der Forschung: Qualitätskontrolle der Inhaltsanalyse (Reliabilität und Validität)	2
2.2	Reliabilitätstest	3
2.3	Relevanz des Inter-Codierer-Reliabilitätstests	3
3	Wie? Vorgehensweise, Methode(n) und Lösungsansätze	4
3.1	Inhaltlicher Auftrag: Operationalisierung auf Statement-Ebene	4
3.2	Methodischer Auftrag: Vorbereitung und Koordination des Inter-Codierer-Reliabilitätstest	4
3.2.1	Berechnung des Reliabilitätskoeffizienten	5
3.2.2	Ergebnisdarstellung und Beurteilung	5
3.2.3	Konkreter Ablauf in diesem Projekt:	6
4	Wohin? Zielsetzung	6
4.1	Inhaltlicher Auftrag: Operationalisierung auf Statement-Ebene	6
4.2	Methodischer Auftrag: Vorbereitung und Koordination des Inter-Codierer-Reliabilitätstests	7
5	Wann? Zeitplan, Schnittstellen, Abhängigkeiten	7
5.1	Arbeitsschritte und Zeitplanung	7
5.2	Schnittstellen und Abhängigkeiten	7
5.3	Risikoanalyse	8
6	Offene Fragen für Diskussion der Version 1 des Konzepts mit C. Dalmus	8

Abbildungsverzeichnis

1	Matrix für Codierübereinstimmungen bei einer Variable (Früh, 2017, S. 184) .	5
---	--	---

1 Was? Thema und Fragestellung

Die Arbeitsgruppe (AG) 5 hat zwei Aufträge zugewiesen erhalten:

1. **Inhaltlicher Auftrag:** Operationalisierung der theoretischen Kategorien (Leistungsindikatoren) auf Statement-Ebene
2. **Methodischer Auftrag:** Vorbereitung und Koordination des Inter-Codierer-Reliabilitätstest zu V1 und V2 des Codebuchs

Mit der Ausführung des inhaltlichen Auftrags wird ein Beitrag an das Codebuch geleistet, indem die Kategorien für die Analyseeinheit «Statements» operationalisiert werden. Erst dadurch werden diese messbar und können als Basis für die Beantwortung der Forschungsfrage dienen: **Wie kann die Leistungsfähigkeit von unterschiedlichen Anbietern (Legacy Media und neue Online-Only-Medien) mit Hilfe einer Inhaltsanalyse und dazugehörigen Leistungsindikatoren erfasst werden?**

Die methodische Aufgabe der AG 5 liegt darin, ein Konzept für die Durchführung von Reliabilitätstests zu erstellen und im Rahmen des Projekts auch konkret einen solchen zu vollziehen. Trotz der Formulierung unseres Auftrags durch die Dozierenden, welche die Durchführung von mehreren Tests suggeriert, geht sowohl die AG 5 als auch die Projektleitung aktuell davon aus, dass nur ein Prüfdurchgang geplant ist. Nach Früh (2017, S. 185) ist dann ein zweiter Test nötig, wenn die Mängel (besonderes bei zentralen Kategorien) gross sind. Ansonsten kann mit der Codierung direkt nach den ersten Korrekturen begonnen werden. Der Reliabilitätstest hat den Zweck, die Schwachstellen der ersten Version des Codebuchs zu identifizieren, um jene Erkenntnisse anschliessend an die AG 4 für die Überarbeitung des Codebuchs zur zweiten, finalen Version weiterzugeben. Ausserdem könnten einzelne Codierer*innen identifiziert werden, die weitere Schulungen zum Codieren benötigen. Die im Rahmen des Tests zu untersuchenden Datensätze werden von der AG 2 gesammelt und den Codierenden zur Verfügung gestellt; für jede Person ist eine Gesamtzahl von 20 Artikeln vorgesehen.

2 Warum? Stand der Forschung und Relevanz des Inter-Codierer-Reliabilitätstest

Nachstehend wird nur auf die Relevanz des methodischen Auftrags – die Vorbereitung und Koordination des Reliabilitätstests – eingegangen. Die Methode und Relevanz der Erstellung des Codebuchs und der Erarbeitung eines Leitfadens für die Codierenden fällt in den Zuständigkeitsbereich der Arbeitsgruppe 4.

2.1 Stand der Forschung: Qualitätskontrolle der Inhaltsanalyse (Reliabilität und Validität)

Nach Rössler (2017, S. 205) hängt die Aussagekraft einer Inhaltsanalyse im Wesentlichen von der Sorgfalt ab, mit der die Forschungsfrage in die Untersuchungsanlage übersetzt und wie das Analyseinstrument angewendet wurde. Ob Instrument und Messung tatsächlich geeignet sind, um Antworten auf die Forschungsfrage zu finden, ist eine Qualitätsfrage und bedarf einer Kontrolle.

Die wichtigsten Gütekriterien für Inhaltsanalysen sind deshalb die **Reliabilität** (auch Zuverlässigkeit) und die **Validität** (auch Gültigkeit) (Brosius, Haas & Koschel, 2016, S. 51 f.):

- **Reliabilität des Messinstruments:** Zuverlässigkeit der Messung und somit auch ihrer Wiederholbarkeit (bei wiederholter Messung sollte das gleiche Ergebnis erzielt werden).
- **Validität der Erhebung:** Gültigkeit der Messung (sie gibt an, ob ein Instrument tatsächlich das misst, was es messen soll).

Gemäss Rössler (2017, S. 206) kann die Reliabilität durch den Vergleich von unterschiedlichen Messungen mit dem gleichen Instrument bestimmt werden und das Ausmass der Abweichung lässt sich dann (ziemlich) genau mit dem Reliabilitätskoeffizienten angeben. Rössler betont, dass hingegen die Validität einer Inhaltsanalyse einer unabhängigen Messung bedarf und darüber hinaus oft auch einer inhaltlichen Diskussion.

Im Rahmen dieses Projekts wird nur die Reliabilität des Messinstruments überprüft mit Hilfe des Inter-Codierer-Reliabilitätstests.

2.2 Reliabilitätstest

Der Reliabilitätstest gehört im Ablauf der hier angewendeten Forschungsmethode der Inhaltsanalyse zur Phase der «Systematischen Verdichtung»: Nach der Operationalisierung der zentralen Konzepte und dem Erstellen des Codebuchs, muss dieses auf seine Zuverlässigkeit und somit auch auf die Wiederholbarkeit der Messung, unabhängig von der codierenden Person, überprüft werden (Dahinden & Dalmus, 2020, S. 19).

Nach Rössler (2017, S. 207) gibt es drei verschiedene Typen der Reliabilitätsmessung, die hier – um ein gemeinsames Verständnis der Begriffe zu erhalten – kurz definiert werden:

- **Intra-Codier-Reliabilität** (bei Rössler «Intracoder»): Wenn der gleiche Codierer am Ende der Analyse nochmals das gleiche Material codiert, ohne seine ersten Codierungen «zu kennen». Diese wird bei einer längeren Projektdauer mit der Inter-Codierer-Reliabilität kombiniert (Früh, 2017, S. 179).
- **Inter-Codierer-Reliabilität:** Wenn zwei Personen das gleiche Material getrennt voneinander codieren und dann verglichen wird, ob die jeweiligen Objekte den gleichen Kategorien zugeordnet wurden.
- **Forscher-Codierer-Reliabilität:** Wenn der Forscher selber codiert und die Resultate mit denen der unabhängigen Codierer verglichen werden.

2.3 Relevanz des Inter-Codierer-Reliabilitätstests

Wenn nach der Durchführung des Inter-Codierer-Reliabilitätstests die Ergebnisse nicht innerhalb eines bestimmten, festgelegten Grades übereinstimmen, muss untersucht werden, wo die Schwachstellen des Codebuchs und der Codieranweisungen liegen und wie diese zu bereinigen sind. Problematische Kategorien können so identifiziert, das Kategoriensystem angepasst und die Anweisungen für die Codierung verbessert werden.

Nach Früh (2017, S. 179) ist der Inter-Codierer-Reliabilitätstest unverzichtbar, da die Reliabilität des Messinstruments für die Objektivität der Inhaltsanalyse notwendig ist. Hingegen wird die Intra-Codier-Reliabilität nur bei Bedarf ermittelt.

3 Wie? Vorgehensweise, Methode(n) und Lösungsansätze

3.1 Inhaltlicher Auftrag: Operationalisierung auf Statement-Ebene

Die Operationalisierung bedingt vorerst die Definition der Kategorien im Codebuch (Leistungsindikatoren). Eine Empfehlung derer wird erst noch von der Projektleitung (PL) vorgelegt. Nichtsdestotrotz wurde mit ihr vereinbart, dass erste Überlegungen und Anhaltspunkte gesammelt und in die vorliegende erste Version des Konzepts einfließen.

Um herauszufinden, ob die Medienangebote die funktionalen und normativ-demokratieorientierten Qualitätskriterien nach ersten Angaben zum Projekt Dahinden und Dalmus (2020, S. 32) und in Anlehnung an Rössler (2017, S. 44–45) erfüllen, soll auf Statement-Ebene ...

- das Thema, dessen Aktualität und die Art und Weise, wie die Thematik dargestellt wird,
- eine allfällige Positionierung und Argumentation und
- die verwendete Sprache und vorzufindende Rhetorik ...

betrachtet werden. Für diese (und evtl. weitere) Kategorien müssen dann in einem zweiten Schritt Indikatoren (oder auch: Ausprägungen) gefunden werden, anhand derer eine Messung möglich ist. So könnte man bspw. für die Kategorie «Argumentation» die Indikatoren «Aussage wird mit Fakten unterstrichen» oder «Aussage entspricht einer subjektiven Wahrnehmung» aufführen. Schliesslich muss für jeden Indikator das geeignete Skalenniveau eruiert werden. Um Missverständnissen und Unklarheiten vorzubeugen, werden diese Einträge um Definitionen und Beispiele ergänzt.

3.2 Methodischer Auftrag: Vorbereitung und Koordination des Inter-Codierer-Reliabilitätstest

Gemäss Rössler (2017, S. 207) werden jeweils die Codierungen von zwei Codierern paarweise miteinander verglichen. Dabei gibt es für diese Prüfung anscheinend in den gängigen Statistikprogrammen keine Auswertungsprozesse, sodass die Berechnung mit eigenen Excel-Sheets o. Ä. programmiert werden, oder von Hand ausgerechnet werden muss.

Rössler (2017, S. 208) stellt als Basis für die konkrete Durchführung des Inter-Codierer-Reliabilitätstests die folgenden drei Fragen:

1. Was gilt als Übereinstimmung und was als Abweichung der Werte?
2. Wie berechnet sich der Grad der Übereinstimmung?
3. Wie viel Material muss für den Test bearbeitet werden?

3.2.1 Berechnung des Reliabilitätskoeffizienten

Zur Berechnung des Reliabilitätskoeffizienten empfiehlt Rössler (2017, S. 212) das – trotz seinen Schwächen – weit verbreitete klassische, einfach zu ermittelnde «*Überschneidungsmass nach Holsti*». Mit diesem Mass wird die Zahl der übereinstimmenden Codierungen zur Gesamtzahl aller Codierungen von zwei Personen in Beziehung gesetzt. In der Formel wird die Zahl der gemeinsamen Codierungen zweier Codierer*innen verdoppelt und anschliessend durch die Summe aller von beiden gesamt vorgenommenen Codierungen geteilt. Die resultierenden Werte liegen dann als Koeffizient zwischen 0 und 1 und können somit als Übereinstimmung zwischen 0 und 100 Prozent ausgedrückt werden. Diese Formel muss für jede Kategorie einzeln und für jedes Paar von Codierer*innen angewendet werden:

Codier-Übereinstimmung nach Holsti:

$$C_R = 2 \cdot \frac{C_{\ddot{U}}}{(C_A + C_B)} \quad (1)$$

- C_R = Reliabilitätskoeffizient der Codierung
- $C_{\ddot{U}}$ = Zahl übereinstimmender Codierungen
- C_A = Zahl der Codierungen von Codierer*in A
- C_B = Zahl der Codierungen von Codierer*in B

3.2.2 Ergebnisdarstellung und Beurteilung

Nachdem für je zwei Codierer*innen deren Übereinstimmung bezüglich jeder Kategorie (mithilfe der Formel nach Holsti) ermittelt wurde, lässt sich das Ergebnis des Inter-Codierer-Reliabilitätstests anschliessend auf drei Ebenen analysieren:

1. **Nach Kategorien:** Mittelwert aller Koeffizienten pro Kategorie (über alle Codiervergleiche).
2. **Nach Codierer*innen:** Mittelwert aller Koeffizienten für eine*n Codierer*in (über alle Kategorien).
3. **Als Matrix:** Übereinstimmungen nach Kategorien und Codierer*innen.

	Codierer 1	Codierer 2	Codierer 3	Codierer 4	Mittelwert
Codierer 1		.64	.90	.86	.80
Codierer 2			.62	.58	.60
Codierer 3				.85	
Codierer 4			.85		
Mittelwert			.79	.76	.74

Abbildung 1: Matrix für Codierübereinstimmungen bei einer Variable (Früh, 2017, S. 184)

Reliabilitäten sollten gemäss Rössler (2017, S. 210) für die formalen und für die wichtigen inhaltlichen Kategorien getrennt aufgeführt werden. Weil bei formalen Kategorien eine hundertprozentige Übereinstimmung eigentlich schon fast Pflicht ist, aber bei den schwierigen inhaltlichen Kategorien auch geringere Werte zufriedenstellend sind (tendenziell sind für inhaltliche Kategorien Werte ab 0.8 und für formale Kategorien Werte nahe an 1.0 zu fordern). Mögliche Richtwerte und deren Interpretation (für den Koeffizienten Cohens Kappa):

Wert	Interpretation
0 – .20	–
.21 – .39	minimal
.40 – .59	schwach
.60 – .79	moderat
.80 – .90	stark
.91	perfekt

3.2.3 Konkreter Ablauf in diesem Projekt:

1. Start mit einem Beispiel für alle Codierer*innen: 1–2 Artikel gemeinsam codieren. Dies könnte von der AG 4 im Rahmen einer Schulung durchgeführt werden. Gemäss Früh (2017, S. 179) soll die Schulung so lange durchgeführt werden, bis alle Beteiligten das Gefühl haben, die Anweisungen zu beherrschen. Innerhalb der Schulung sollen auch Fragen geklärt werden und die Unklarheiten direkt im Codebuch angepasst werden.
2. Es soll für jede Variable definiert werden, was als Übereinstimmung gilt.
3. Alle Projektmitglieder codieren am selben Textmaterial: in unserem Fall codiert jede*r 20 Artikel unabhängig voneinander. Das heisst, ohne sich gegenseitig auszutauschen.
4. Durchführung Inter-Codierer-Reliabilitätstest.
5. Codebuch und Codieranweisung optimieren durch AG 4 (Erstellung V2 des Codebuchs).

4 Wohin? Zielsetzung

4.1 Inhaltlicher Auftrag: Operationalisierung auf Statement-Ebene

In Anlehnung an die von Rössler (2017, S. 105) formulierten Anforderungen an Kategorien sollte unsere Auswahl für die Analyseeinheit «Statements» inkl. deren Ausprägungen trennscharf und vollständig sein. Dadurch soll verhindert werden, dass mit der Verwendung dieses Instruments fehlerhafte Messungen generiert werden. Des Weiteren sollen treffende Definitionen sowie gut gewählte Beispiele dazu führen, dass zwischen den Codierer*innen eine hohe Übereinstimmung und somit eine objektive Genauigkeit des Codebuchs erreicht wird. Wir gehen davon aus: Je präziser und umfassender die Kriterien im Codebuch V1 vorliegen, desto besser sind die Ergebnisse des Reliabilitätstests zu erwarten – und damit einhergehend; desto kleiner fällt der Aufwand für die Überarbeitung des Buchs für die AG 4 aus.

4.2 Methodischer Auftrag: Vorbereitung und Koordination des Inter-Codierer-Reliabilitätstests

Grundsätzlich hoffen wir, dass die gewählte Methode für die Prüfung der Datensätze in Verbindung mit dem Codebuch V1 erfolgreich angewendet werden kann. Der durch uns zu entwickelnde und durchzuführende Test soll anhand des begutachteten Materials aufzeigen können, welche Kategorien und Variablen im Codebuch V1 keine genügend hohe Übereinstimmung aufweisen. Dadurch wird einerseits ermöglicht, das Potential des erarbeiteten Messinstruments zu bewerten und andererseits jene Massnahmen zu erkennen, die zu einer optimierten Version des Codebuchs führen. Schliesslich wäre es wünschenswert, dass das hier erprobte Vorgehen – sofern es sich bewährt – auch im SNF-Projekt Anwendung finden kann.

5 Wann? Zeitplan, Schnittstellen, Abhängigkeiten

5.1 Arbeitsschritte und Zeitplanung

Der Zeitplan wird durch die PL erstellt und muss zwingend von allen AGs als verbindlich angesehen werden. Er hängt einerseits stark davon ab, wie lange die AG 2 und 3 für die Datensammlung und das Datenmanagement benötigen und andererseits wie viel Zeit der AG 4 gegeben wird, um die finale Version 2 des Codebuchs zu erstellen. Wir plädieren stark dafür, genug Zeit für die Durchführung des Tests sowie die anschliessende Bereinigung und Optimierung des Codebuchs in seiner finalen Version einzukalkulieren.

5.2 Schnittstellen und Abhängigkeiten

Ohne die Konzepte der anderen AGs zu kennen, können vorerst folgende Schnittstellen und Abhängigkeiten eruiert werden:

- **AG 1 (Projektleitung):** Die Ausarbeitung und Durchführung der Operationalisierung kann erst nach Fertigstellung der Empfehlung für die Auswahl der Kategorien für das Codebuch begonnen werden.
- **AG 1-5:** Sobald das Set an theoretischen Leistungsindikatoren vorliegt, können die AG 2–5 beginnen, diese Begriffe zu operationalisieren und geeignete Indikatoren, Definitionen und Beispiele suchen. Obschon die Ebene der Statements die unterste Hierarchieebene darstellt, muss dabei nicht erst die Fertigstellung der Kategorien der Ebenen Website, Artikel und Sprecher*in abgewartet werden. Gleichwohl muss darauf geachtet werden, dass das Codebuch in sich stimmig ist – es bildet somit als Ganzes eine konzeptuelle Schnittstelle, die stets auch als Gesamtheit betrachtet werden muss.
- **AG 2 (Rohdatensammlung):** Das Sammeln der Beiträge für die Inhaltsanalyse ist Basis für das Durchführen der Codierarbeit.
- **AG 4 (Codebuchentwicklung):** Um den Inter-Codierer-Reliabilitätstest machen zu können, muss die erste Version des Codebuchs vorliegen.
- **AG 3-5 (die Operationalisierenden):** Jedes Mitglied dieser vier Arbeitsgruppen muss die Codierarbeit der insgesamt 20 Artikel fristgerecht erledigen, damit der Test durchgeführt werden kann.

Weitere Schnittstellen und Abhängigkeiten werden sich nach der Vorstellung der Konzepte der einzelnen Arbeitsgruppen zeigen.

5.3 Risikoanalyse

Das Projekt lebt von der Mitarbeit und dem Engagement von knapp 20 Personen, was personelle Risikofaktoren mit sich bringt. Wie die grob skizzierten Abhängigkeiten zeigen, würde ein nicht eingehaltener Zeitplan einer anderen Arbeitsgruppe die Erledigung der eigenen Aufgaben stark beeinträchtigen. Verstärkt wird das für unsere Arbeitsgruppe insbesondere dadurch, weil sich jene Arbeitsschritte, die wir koordinieren, eher am Ende des Projekts befinden. Zudem besteht das Risiko, dass Lieferobjekte, die nicht den Anforderungen entsprechen, die Qualität der weiteren Arbeit damit unumgänglich beeinflussen. Auch zu erwähnen ist die Schnittstelle 'Codebuch', die das Mitwirken aller Projektmitglieder voraussetzt und viel Koordination und Kontrolle verlangt.

6 Offene Fragen für Diskussion der Version 1 des Konzepts mit C. Dalmus

- **Anzahl der Reliabilitätstests:** Die Formulierungen der Aufträge an die AG 4 und AG 5 sind unseres Erachtens widersprüchlich. Sind ein (zu Codebuch V1) oder zwei Tests (zu Codebuch V1 und V2) vorgesehen?
- **Codebuch:** Handelt es sich bei V2 um das finale Produkt?
- **Codierauftrag:** Jede Person aus AG 2-5 codiert 20 Artikel (würde gesamthaft 320 codierte Artikel ergeben), haben wir das richtig verstanden? In der Beschreibung heisst es "jede Person codiert insgesamt 20 Artikel" – fließen alle 320 Artikel in den Test mitein oder nur ein Bruchteil davon bzw. wird nur vor dem Test codiert oder nochmals danach?
- **Inter-Codierer-Reliabilitätstest:** Wie läuft so ein Test konkret ab? (Auswahl richtige Formel, Berechnung von Hand oder mit Software?, evt. Beispiel aus der Praxis zeigen)
- **Inter-Codierer-Test vs. Forscher-Codierer-Test:** Die Forschenden selbst sollten sich nach Früh (2017, S. 179) auch an dem Test beteiligen. Wenn die Projektgruppe selber operationalisiert und dann codiert, handelt es sich dann um welche der beiden Varianten?
- **Interpretation:** Wer interpretiert die Ergebnisse des Tests – AG 4 oder AG 5?

Literatur

- Brosius, H.-B., Haas, A. & Koschel, F. (2016). Inhaltsanalyse I: Grundlagen. In H.-B. Brosius, A. Haas & F. Koschel (Hrsg.), *Methoden der empirischen Kommunikationsforschung: Eine Einführung* (S. 137–152). Studienbücher zur Kommunikations- und Medienwissenschaft. doi:10.1007/978-3-531-19996-2_8
- Dahinden, U. & Dalmus, C. (2020, September). Projektkurs Lokaljournalismus - Folien Unterricht. FH Graubünden.
- Früh, W. (2017). *Inhaltsanalyse* (9. Aufl.). Konstanz, München: UVK.
- Rössler, P. (2017). *Inhaltsanalyse* (3., völlig überarbeitete Auflage). UTB basics. Konstanz: UVK.