

Berechnung von Reliabilitätskoeffizienten

Nachfolgend wird die Berechnung von drei verschiedenen Reliabilitätskoeffizienten beschrieben und erläutert:

- Übereinstimmungskoeffizient (r_U)
- Reliabilitätskoeffizient nach Holsti (r_H)
- Krippendorff's α (r_α)

Will man einen Reliabilitätstest durchführen, gilt es als erstes, eine geeignete Form zu finden, um die übereinstimmenden und nicht übereinstimmenden Codierentscheidungen zu dokumentieren, um anschließend auf Basis dieser Dokumentation einen Reliabilitätskoeffizienten berechnen zu können. Die Dokumentation sollte in Tabellenform und variablenweise erfolgen. In die erste Spalte einer solchen Tabelle sollten zeilenweise die Nummern der Untersuchungseinheiten eingetragen werden, die für den Reliabilitätstest codiert wurden. Diese Nummerierung ist sehr wichtig, denn ohne eine exakte Zuordnung mehrerer Codierungen zur selben Untersuchungseinheit ist die Berechnung eines Reliabilitätskoeffizienten nicht möglich. Die weiteren Spalten dieser Tabelle sollten in Zahlenform die Entscheidungen der einzelnen Codierer enthalten (vgl. Tab. 7).

Tabelle 7: Ausschnitt aus dem Reliabilitätstest der Beispielstudie

Nummer der Untersuchungseinheit	Codierer 1	Codierer 2	Codierer 3	Übereinstimmung
1	0	0	0	Ja
2	0	0	0	Ja
3	2	1	0	Nein
4	0	0	1	Nein
5	2	2	2	Ja
6	0	0	0	Ja

Tabelle 7 veranschaulicht das oben beschriebene Dokumentationsprinzip am Beispiel von drei Codierern, die im Rahmen des Reliabilitätstests der Beispielstudie sechs Untersuchungseinheiten codiert haben. Die dargestellten Codierentscheidungen beziehen sich auf die Variable „Betroffenheit/Reichweite“ (N10). Die Variable N10 ist Teil der Erfassung der Nachrichtenfaktoren und sieht drei mögliche Variablenausprägungen vor.²

Es ist – wie schon bei den vielen Möglichkeiten der Stichprobenziehung – angesichts einer Vielzahl von Reliabilitätskoeffizienten gar nicht so einfach herauszufinden, welcher davon eigentlich im Sinne einer „good-practice“ geeignet ist.³ Nachfolgend ist deshalb ausführlich erklärt, wie man verschiedene Reliabilitätskoeffizienten berechnet.

Die letzte Spalte von Tabelle 7 zeigt an, dass die drei Codierer in ihren Codierentscheidungen vier von sechs Mal global übereinstimmt haben. Zur Berechnung des **Übereinstimmungskoeffizienten** setzt man die Zahl aller komplett übereinstimmenden Codierungen mit der Gesamtzahl aller Codierung ins Verhältnis. Für die Werte in Tabelle 7 ergibt dies dann einen Reliabilitätswert $r_U = 4/6 = .67$. Dieser Wert besagt, dass der Übereinstimmungsgrad dieser drei Codierer bei 67% liegt.

Was sagt ein Reliabilitätskoeffizient eigentlich aus? Wie ist er zu interpretieren? Perfekte Reliabilität entspricht einer Übereinstimmung von 100% ($r = 1$). Neuendorf (2002: 143) kommt nach Sichtung der Methodenliteratur zu dem Schluss, dass man eine Übereinstimmung von über 90% ($r \geq .90$) als sehr gut und eine Übereinstimmung von über 80% ($r \geq .80$) zumindest als noch gut bezeichnen kann. Welcher Schwellenwert unterhalb einer Übereinstimmung von 80% ($r < .80$) allerdings noch akzeptabel ist, darüber herrscht in der Methodenliteratur offener Streit (vgl. Lombard/Snyder-Duch/Bracken 2002).

Es empfiehlt sich, bei der Interpretation von Reliabilitätswerten die in der Methodenliteratur empfohlenen Schwellenwerte stets in Beziehung zum Schwierigkeitsgrad einer Variable zu setzen. Rössler (2005: 192)

² Die Ausprägungen der Variable N10 lauten: „0 = Nicht genannt“, „1 = Geringer Betroffenheitskreis“, „2 = Großer Betroffenheitskreis“.

³ Einen gut verständlichen „state-of-the-art“-Überblick zu den Stärken und Schwächen verschiedener Reliabilitätskoeffizienten findet man bei Krippendorff (2004a).

argumentiert, dass die Reliabilitätswerte von technischen oder formalen Variablen (z.B. Datum eines Artikels usw.) eigentlich nahezu perfekt reliabel codiert werden müssten.

Er empfiehlt eine etwas anspruchsvollere Methode der Reliabilitätsberechnung: den **Holsti-Reliabilitätskoeffizienten** (vgl. Rössler 2005: 190). Dieser beruht auf der Logik des Paarvergleichs. Inspiziert man Tabelle 7 noch einmal genauer, fällt bei einem paarweisen Vergleich in Zeile vier und fünf auf, dass Codierer 1 und 2 übereinstimmend gearbeitet haben.

Das Reliabilitätsprogramm PRAM erleichtert die Durchführung eines Reliabilitätstests wie dem von Holsti, denn es berechnet automatisch alle paarweisen Vergleiche und den entsprechenden Durchschnittswert, wenn man die Werte aus Tabelle 7 – wie in Abbildung 2 dargestellt – neu anordnet.⁴

Abbildung 2: Neuordnung der Codierentscheidungen in einem Tabellenkalkulationsprogramm

⁴ Das Programm PRAM (Version 0.4.5) und seine Bedienungsanleitung können kostenlos heruntergeladen werden: <http://www.geocities.com/skymegsoftware/pram.html>, abgerufen am 23.02.2009.

	A	B	C
1	Coder	Unit	N10
2	1	1	0
3	2	1	0
4	3	1	0
5	1	2	0
6	2	2	0
7	3	2	0
8	1	3	2
9	2	3	1
10	3	3	0
11	1	4	0
12	2	4	0
13	3	4	1
14	1	5	2
15	2	5	2
16	3	5	2
17	1	6	0
18	2	6	0
19	3	6	0

Die Berechnung des Holsti-Koeffizienten erfolgt zweistufig: Zuerst berechnet man alle möglichen Paarvergleiche und bestimmt anschließend den Durchschnitt aller möglichen Paarvergleiche (vgl. Abb. 3).⁵

Abbildung 3: PRAM-Ergebnisdarstellung

	A	B	C
1	Holsti's Coefficient of Reliability		
2			
3	Coder Pair		
4	1,2		0,833
5	1,3		0,667
6	2,3		0,667
7			
8	Average		0,722

⁵ Man beachte dabei jedoch: Bei einer ansteigenden Zahl von Codierern wächst die Zahl der zu vergleichenden Paare nicht-linear an (vgl. Kolb 2004). Tabelle 7 beinhaltet zum Beispiel drei Vergleichspaare (fünf Codierer ergäben bereits zehn Vergleichspaare, zehn Codierer sogar 45 Vergleichspaare usw.).

Als paarweise Übereinstimmung zwischen Codierer 1 und Codierer 2 berechnet PRAM einen Reliabilitätskoeffizienten von $r_{12} = .83$, zwischen Codierer 1 und Codierer 3 einen Reliabilitätswert von $r_{13} = .67$ und zwischen Codierer 2 und Codierer 3 einen Reliabilitätswert von $r_{23} = .67$. PRAM errechnet außerdem die durchschnittliche Reliabilität nach Holsti $r_H = .72$.

Vergleicht man beide Formen der Reliabilitätsberechnung, dann fällt der Übereinstimmungskoeffizient mit $r_U = .67$ in der Beurteilung der Reliabilität etwas konservativer aus als der Holsti-Koeffizient mit $r_H = .72$. Wieso kommen beide Koeffizienten nicht zum gleichen Ergebnis?

Die Berechnungsweise des Holsti-Koeffizienten berücksichtigt, dass Codierer 1 und Codierer 2 in ihren Codierungen fast immer übereinstimmen und dass es vor allem Codierer 3 ist, der abweichend codiert. Kurz gesagt: Während die Berechnung des Übereinstimmungskoeffizienten nach dem Prinzip des schwächsten Glieds der Kette erfolgt, ist das bei der etwas liberaleren Berechnungsmethode nach Holsti nicht der Fall.

Obwohl für die Forschungspraxis empfohlen (vgl. Rössler 2005: 190), wird die Verwendung des Holsti-Reliabilitätskoeffizienten unter Anführung verschiedener Argumente zuweilen heftig kritisiert. Die Argumente lauten (vgl. zusammenfassend Hayes/Krippendorff 2007):

- **Zahl der Variablenausprägungen:** Die Höhe des Holsti-Koeffizienten wird beeinflusst von der Zahl der Ausprägungen, die eine Variable aufweist.
- **Skalenniveau:** Die Berechnung des Holsti-Koeffizienten ist nur für nominalskalierte Variablen zulässig.
- **Fehlende Werte:** Die Berechnung des Holsti-Koeffizienten setzt voraus, dass die Werte aller Codierer komplett vorliegen.
- **Präzision des Koeffizienten:** Eine Berechnung von Konfidenzintervallen und damit die Angabe der Genauigkeit des Koeffizienten sind für den Holsti-Koeffizienten nicht möglich.
- **Fehlende Korrektur für zufällige Übereinstimmungen:** Die Berechnung des Holsti-Koeffizienten sieht keine Korrektur für zufällige Übereinstimmungen von Codierern beim Reliabilitätstest vor.

Angesichts dieser ernstzunehmenden Kritik an der Holsti-Methode – und insbesondere wegen der fehlenden Zufallskorrektur – ist ein alternativer Reliabilitätskoeffizient entwickelt worden: **Krippendorff's α** (vgl.

Krippendorff 2004a; Krippendorff 2004b; Hayes/Krippendorff 2007; Krippendorff 2008).

Krippendorff's α lässt sich mithilfe des Statistikprogramms SPSS und einer zusätzlichen Macro-Datei berechnen. Die Berechnung von Krippendorff's α ist sehr ausführlich im Internet unter der Adresse <http://www.comm.ohio-state.edu/ahayes/SPSS%20programs/kalpha.htm> dokumentiert:

1. Öffne in SPSS ein neues Datenfenster.
2. Klicke auf den untenstehenden Reiter „Variable View“ (Variablenansicht) und trage in die Spalte „Name“ zeilenweisen die Namen der Codierer ein (vgl. Abb. 4). Im SPSS-Makro werden die Codierer als „obs“ (Abkürzung für „Observer“) bezeichnet. Codierer 1 bezeichnet man gemäß der SPSS-Makro-Konvention als „obsa“ – die Kürzel stehen für „Observer A“, „Observer B“, „Observer C“ usw. Bei 14 Codierern würde man die Namen „obsa“ bis „obsn“ vergeben.
3. Klicke nun auf den untenstehenden Reiter „Data View“ (Datenansicht) und trage die einzelnen Werte der Codierer ein, die pro Untersuchungseinheit beim Reliabilitätstest vergeben wurden. Abbildung 5 zeigt einen Ausschnitt der in SPSS übertragenen Werte aus Tabelle 7.
4. Lade das SPSS-Skript „kalphav2_1.SPS“ herunter: http://www.comm.ohio-state.edu/ahayes/SPSS%20programs/kalphav2_1.SPS
5. Klicke in der oberen Menüleiste auf „File“ (Datei), klicke dann auf „Open“ (Öffnen) und dann auf „Syntax“. Öffne die soeben abgespeicherte SPSS-Marko-Datei „kalphav2_1.SPS“ im SPSS-Syntax-Fenster, markiere dann den gesamten Text der Datei und drücke anschließend die Tastenkombination STRG+r. Das Drücken dieser Tastenkombination aktiviert das SPSS-Makro.
6. Um Krippendorff's α für nominalskalierte Variablen zu berechnen, ergänze die SPSS-Marko-Datei um folgende Zeile und ersetze die kursiv geschriebenen Worte mit entsprechenden Angaben:

`kalpha judges = Liste mit Codierernamen /level = Skalenniveau/detail = 1/boot = 5000.`

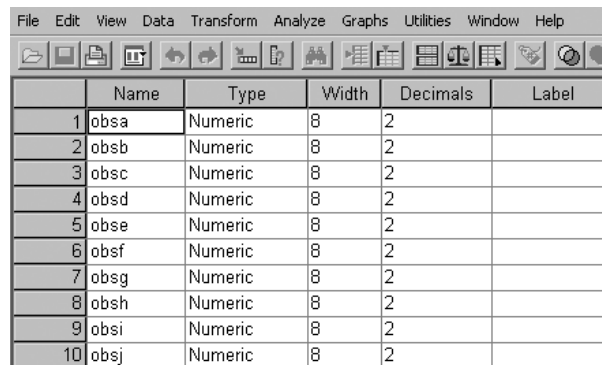
Das SPSS-Makro erlaubt die Berechnung von Krippendorff's α für alle theoretisch möglichen Skalenniveaus. Die Optionen sind: 1 = Nominalskala, 2 = Ordinalskala, 3 = Intervallskala, 4 = Ratioskala. Bei unserer Beispielstudie lautet der SPSS-Befehl dann:

```
kalpha judges = obsa obsb obsc obsd obse obsf obsg obsh obsi obsj obsk obsl obsm  
obsn /level = 1/detail = 1/boot = 5000.
```

Dieser Befehl wird ausgeführt, wenn man diese Zeile im SPSS-Syntax-Fenster markiert und die Berechnung durch Drücken der Tastenkombination STRG+r auslöst.

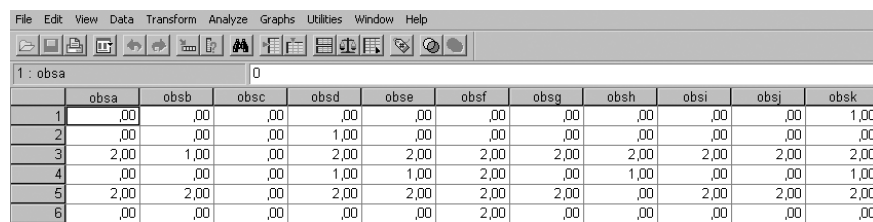
7. Nach der Berechnung von Krippendorfs α öffnet sich ein SPSS-Output-Fenster, das in der Spalte „Alpha“ den berechneten Reliabilitätskoeffizienten enthält.

Abbildung 4: Variablenansicht in SPSS



	Name	Type	Width	Decimals	Label
1	obsa	Numeric	8	2	
2	obsb	Numeric	8	2	
3	obsc	Numeric	8	2	
4	obsd	Numeric	8	2	
5	obse	Numeric	8	2	
6	obsf	Numeric	8	2	
7	obsf	Numeric	8	2	
8	obsh	Numeric	8	2	
9	obsi	Numeric	8	2	
10	obsj	Numeric	8	2	

Abbildung 5: Datenansicht in SPSS



	obsa	obsb	obsc	obsd	obse	obsf	obsf	obsf	obsh	obsi	obsj	obsk
1	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	1,00
2	,00	,00	,00	1,00	,00	,00	,00	,00	,00	,00	,00	,00
3	2,00	1,00	,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00
4	,00	,00	,00	1,00	1,00	2,00	,00	1,00	,00	,00	,00	1,00
5	2,00	2,00	,00	2,00	2,00	2,00	2,00	,00	2,00	2,00	2,00	2,00
6	,00	,00	,00	,00	,00	2,00	,00	,00	,00	,00	,00	,00

Gemäß der Formel zur Berechnung des Übereinstimmungskoeffizienten ergab sich in unserem Beispiel ein Reliabilitätskoeffizient von $r_U = .67$ und gemäß der Holsti-Formel ein Reliabilitätskoeffizient von $r_H = .72$. Krippendorff's α hingegen ergibt ein Reliabilitätswert von $r_\alpha = .47$.

Was sagen die unterschiedlichen Reliabilitätskoeffizienten über die Übereinstimmung der drei Codierer aus? Welchen Reliabilitätswert sollte man zur Beurteilung der Reliabilität heranziehen?

Statistisch gesehen wäre es beste Forschungspraxis, sich bei der Beurteilung der Reliabilität ausschließlich auf Krippendorff's α zu verlassen.⁶ Wieso das? Reliabilitätstests sind statistisch gesehen Schätzungen – und zwar Übereinstimmungsschätzungen. Da man aus Zeit- und Kosten-

⁶ Kolb (2004) argumentiert – wenn auch etwas skeptischer – in ähnlicher Weise.

gründen nur eine begrenzte Zahl von Untersuchungseinheiten beim Reliabilitätstest berücksichtigen kann, ist es unmöglich, die „wahre“ Reliabilität zu bestimmen. Wenn Reliabilitätstests schätzenden Charakter haben, dann sind demzufolge Reliabilitätswerte nichts anderes als Schätzwerte für die „wahre“, aber unbekannte Reliabilität. Die Präzision von Schätzwerten hängt laut statistischer Theorie u.a. von der Stichprobengröße ab. Die statistischen Eigenschaften von Krippendorff's α erlauben die Berechnung eines 95%-Konfidenzintervalls (95%-KI). Das Konfidenzintervall wiederum gibt Auskunft über die Präzision der Reliabilitätsschätzung α .

In Tabelle 8 sind die nach Holsti und Krippendorff berechneten Reliabilitätskoeffizienten gegenübergestellt. Für Krippendorff's α ist zusätzlich das 95%-Konfidenzintervall ausgewiesen. Die Variablen von „Technische Nr. der Untersuchungseinheit“ (T1) bis einschließlich „Bildanalyse“ (B1) wurden von den Codierern der Beispielstudie sehr gut bis gut übereinstimmend codiert: Sowohl die Holsti- als auch die Krippendorff-Koeffizienten liegen über den empfohlenen Schwellenwerten von $\geq .80$ oder sogar $\geq .90$. Die Reliabilitätskoeffizienten der Variable „Thema des Artikels“ (TH1) sowie fast alle Nachrichtenfaktoren zeigen für sich genommen mehr oder weniger große Übereinstimmungsmängel an. Außerdem weichen mitunter die Holsti- und die Krippendorff-Koeffizienten stark voneinander ab.

Nach Holsti ergibt sich beispielsweise für die Variable „Überraschung“ (N2) ein Reliabilitätswert von $r_H = .48$ und nach Krippendorff ein Reliabilitätswert von $r_\alpha = .12$.

Was sind die Gründe für diese nicht nur schlechten, sondern darüber hinaus auch noch unterschiedlichen Reliabilitätsergebnisse? Haben wir Forscher die Codierer im Vorfeld zu wenig geschult? Haben die Codierer trotz Schulung „schlecht“ gearbeitet und die Codierregeln jeweils anders ausgelegt? Haben wir bei der Planung des Reliabilitätstests Fehler gemacht? Oder waren die Nachrichtenfaktorenvariablen schlichtweg zu schwer zu codieren?

Tabelle 8: Ergebnis des Reliabilitätstests der Beispielstudie

Variable	Variablenblock	Reliabilitätskoeffizient	
	Variablenname	Holsti	Krippendorfs α [95%-KI]
t1	Techn. Nr. der UE	.98	.97 [.96–.99]
	<i>Datum</i>		
t3a	Tag	1.00	1.00
t3b	Monat	.98	.97 [.95–.98]
t3c	Jahr	1.00	1.00
t4	Untersuchungstyp	1.00	1.00
	<i>Pressemitteilung</i>		
pf2	Wortanzahl	1.00	1.00
pf3	Form der Pressemitteilung	.88	.81 [.76–.86]
	<i>Zeitungsartikel</i>		
zf2	Wortanzahl	1.00	1.00
zf3	Titel der Zeitung	1.00	1.00
zf4	Paginierung	1.00	1.00
zf5	Rubrik/Ressort	1.00	1.00
zf6	Genre/Journalistische	1.00	1.00
	Form/Stilform		
zf7a	Urheber des Artikels	.96	.95 [.91–.97]
	<i>Agenturbeitrag</i>		
af2	Wortanzahl	1.00	1.00
af3	Agenturdienst	1.00	1.00
af4	Veranstaltungshinweis	1.00	1.00
	<i>Quellenanalyse</i>		
üa1	Explizite Quellennennung		
u1	Link-Angabe	1.00	1.00
	<i>Bildanalyse</i>		
ba1	Grafik und/oder Foto vorhanden	.94	.88 [.84–.92]
	<i>Themenanalyse</i>		
ta1	Thema der Untersuchungseinheit	.58	.50 [.45–.55]
	<i>Nachrichtenfaktoren</i>		
nf1	Konflikt/Kontroverse/Problem	.87	.72 [.66–.78]
nf2	Überraschung	.48	.12 [.05–.19]
nf3	Kuriosität	.90	.08 [–.16–.29]
nf4	Aktualität	.78	.63 [.56–.68]
nf5	Nutzen/Erfolg	.78	.36 [.26–.46]
nf6	Schaden/Misserfolg	.79	.16 [.02–.29]
nf7	Personalisierung	.67	.14 [.04–.24]
nf8	Prominenz	.98	.00 [–.54–.46]
nf9	Räumliche Nähe zum Rezipienten	.73	.14 [.03–.26]
nf10	Betroffenheit/Reichweite	.69	.44 [.38–.51]
nf11a	Gesellschaftliches Mega-Thema	.98	.92 [.88–.96]

Es ist nicht einfach, die genauen Gründe für die schwachen Reliabilitätswerte einer einzelnen Variablen zu benennen (vgl. Tab. 8). Mit den Variablen „Überraschung“ (N2) und „Kuriosität“ (N3) sollten beispielsweise spezielle Eigenschaften von Nachrichten erfasst werden, die dem Codierer trotz Standardisierung einen recht hohen Ermessensspielraum lassen. Schlechte Reliabilitätsergebnisse für die Variablen „Überraschung“ (N2) und „Kuriosität“ (N3) sind allerdings gar keine so große Überraschung: In der Nachrichtenwertstudie von Schulz (1976) wies der Nachrichtenfaktor „Überraschung“ zum Beispiel ebenfalls einen der schwächsten Reliabilitätswerte auf. Die Beurteilung der eigenen Reliabilitätswerte sollte also nicht nur anhand empfohlener Schwellenwerte erfolgen, sondern immer auch (sofern möglich) die Ergebnisse bisheriger empirischer Studien und deren Reliabilitätsdokumentationen berücksichtigen. Der Rekurs auf die Reliabilitätsergebnisse von Schulz (1976) soll selbstverständlich kein Vorwand für schwache Reliabilitätswerte sein.

Tabelle 9 zeigt zum besseren Verständnis der nachfolgenden Argumentation die Codierung der Variable „Kuriosität“ (N3) im Detail.

Tabelle 9: Codierung der Variable „Kuriosität“ (N3)

Nr. der	Codierer													
UE ¹	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	1	0	0	0	0	0	0	1	0
6	0	0	1	0	0	1	0	0	0	0	0	0	0	0

¹ Die Abkürzung UE bedeutet: Untersuchungseinheit.

Die ersten vier Untersuchungseinheiten werden von allen Codierern übereinstimmend codiert. Bei der fünften und sechsten Untersuchungseinheit gibt es einige ganz wenige Abweichungen. Auf den ersten Blick wirkt Tabelle 9 so, als müsste man sich bei der Variable „Kuriosität“ keine

Sorgen um mangelnde Reliabilität machen. Die widersprüchlichen Reliabilitätswerte in Tabelle 8 sind den sehr liberalen statistischen Eigenschaften des Holsti-Koeffizienten bzw. den sehr konservativen statistischen Eigenschaften von Krippendorff's α geschuldet.

Krippendorff (2004b: 236-237) rechtfertigt die konservativen Schätzeigenschaften von α wie folgt: Wenn – wie beim Test der Variable „Kuriosität“ – die Untersuchungseinheiten des Reliabilitätstests einen unzureichenden Variationgrad aufweisen, sagen traditionell berechnete Reliabilitätswerte im Grund gar nichts aus. Wieso das? Ein Gedankenexperiment: Wenn in Tabelle 9 alle Werte null gewesen wären, kann man dann wirklich guten Gewissens sagen, dass die Reliabilität perfekt ist, weil alle Codierer die gleichen Werte codiert haben? Berechnet man für dieses Gedankenexperiment den Holsti-Koeffizient, ergibt sich $r_H = 1$. Krippendorff (2004b: 237) lehnt es allerdings strikt ab, in dieser Situation von perfekter Reliabilität zu sprechen; stattdessen fordert er nachdrücklich, den Reliabilitätstest so zu planen, dass die Untersuchungseinheiten genügend Variation aufweisen (d.h. das in den Untersuchungseinheiten alle möglichen Variablenausprägungen auftauchen).

Was könnten man nun bei Vorbereitung und Durchführung eines Reliabilitätstests besser machen als in der Beispielstudie?

- **Mehr Zeit für Codiererschulungen:** Innerhalb des Seminars war es aus Zeitgründen nicht möglich, die für eine Erzielung hoher Reliabilität notwendigen mehrfachen und mehrstündigen Codiererschulungen anzusetzen. Mit mehr Zeit können in der Regel bessere Reliabilitäten erzielen. Bei schlechten ersten Übereinstimmungswerten können außerdem weitere Codiererschulungen und Probecodierungen für Besserung sorgen.
- **Konkretere Themen bei der Themencodierung:** Die Codierung von Themen (TH1) gestaltet sich im Forschungsalltag immer schwierig – und zwar unabhängig davon, ob die Codierung von unerfahrenen Codierern aus einem universitären Seminar oder von speziell geschulten und erfahrenen Codierern vorgenommen wurde. Das liegt an den überdurchschnittlich großen Interpretationsspielräumen und Mehrdeutigkeiten, die eine Themencodierung – aller Standardisierungs- und Klassifikationsbemühungen sowie Ausführlichkeit von Codeerläuterungen zum Trotz – zwangsläufig mit sich bringt. Scheufele (2004: 84) plädiert deshalb dafür, die Themenvariable so zu konzipieren, dass möglichst viele konkrete Themen als Variablenausprägungen vorkommen (enthält die Themenvariable eines Codeplans viele Variablen-

ausprägungen, wird dies üblicherweise in einer gesonderten Liste dokumentiert, um den Umfang des Codeplans nicht zu sprengen). Die vorgeschlagene Konzeption der Themenvariable hat natürlich auch einen Nachteil: Man wird nach der Codierung nicht umhin kommen, die Themenvariablen zu recodieren – sprich: Einzelthemen zu Themenblöcken zusammenzufassen. Genau genommen verlagert man das Reliabilitätsproblem vom Codierer auf die Person, die für Datenauswertung verantwortlich ist.

- **Wenige Ausprägungen bei „schwierigen“ Variablen:** „Schwierige“ Variablen (z.B. die Variablen zur Erfassung der Nachrichtenfaktoren), die dem Codierer selbständige Wertungen und Einschätzungen abverlangen und mithilfe von Ordinalskalen erfasst werden, sind umso leichter übereinstimmend zu codieren, je geringer die Zahl der Variablenausprägungen ist (Rössler 2005: 152).
- **Ausreichende Zahl von Untersuchungseinheiten beim Reliabilitätstest:** Unser Reliabilitätstest basierte auf nur sechs Untersuchungseinheiten, hier: Artikel. Rössler (2005: 191) berichtet, dass ein inhaltsanalytischer Reliabilitätstest auf wenigstens 30 bis 50 Untersuchungseinheiten basieren sollte. Lombard, Snyder-Duch und Bracken (2002: 601) schlagen alternativ vor, dass der Reliabilitätstest auf rund 10% aller Untersuchungseinheiten einer inhaltsanalytischen Studie basieren sollte. Damit lässt sich die oben genannte Schätzung der Reliabilität präzisieren.

Wie sieht nun eine gute Forschungspraxis der Reliabilitätsberechnung aus? Rössler (2005: 190) und Früh (2007: 194-196) plädieren dafür, für die Berechnung von Reliabilitätskoeffizienten die Holsti-Formel zu benutzen. Insbesondere Früh (2007: 196) argumentiert in diesem Zusammenhang, dass bei kleinen Studien das Verhältnis von Berechnungsaufwand und Leistung für Krippendorff's α im Vergleich zum Holsti-Koeffizienten viel zu groß sei.

Eine gute Forschungspraxis sollte zum einen stets die Berechnung und detaillierte Dokumentation von Reliabilitätskoeffizienten vorsehen, und zum anderen sollte – angesichts der optimalen statistischen Qualitäten – die Berechnung von Krippendorff's α immerhin der Maßstab jeder Reliabilitätsberechnung sein.

So wie die Berechnung von Holsti-Reliabilitätskoeffizienten eine besonders geläufige Forschungspraxis darstellt, ist es aus Platzgründen ebenso üblich, nur den kleinsten, den größten und den durchschnittlichen Reliabilitätskoeffizienten auszuweisen. Lombard Snyder-Duch und Bra-

cken (2002) sowie Krippendorff (2004a) sprechen sich gegen diese Praxis aus und argumentieren, dass ein Reliabilitätsdurchschnitt, in den alle Variablen des Codeplans eingegangen sind, aufgrund der hohen Reliabilität der formalen Variablen zwangsläufig überdurchschnittlich gut sein wird.

Literatur

- Früh, Werner (2007): Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK/UTB.
- Hayes, Andrew F./Krippendorff, Klaus (2007): Answering the Call for a Standard Reliability Measure for Coding Data. In: *Communication Methods and Measures* 1(1), 77–89.
- Kolb, Steffen (2004): Verlässlichkeit von Inhaltsanalysedaten – Reliabilitätstest, Errechnen und Interpretieren von Reliabilitätskoeffizienten für mehr als zwei Codierer. In: *Medien & Kommunikationswissenschaft* 52(3), 335–354.
- Krippendorff, Klaus (2004a): Reliability in Content Analysis. Some common Misconceptions and Recommendations. In: *Human Communication Research* 30(3), 411–433.
- Krippendorff, Klaus (2004b): *Content Analysis. An Introduction to Its Methodology*. Thousand Oaks u.a.: Sage.
- Lombard, Matthew/Snyder-Duch, Jennifer/Bracken, Cheryl C. (2002): Content analysis in mass communication. Assessment and reporting of intercoder reliability. In: *Human Communication Research* 28(4), 587–604.
- Neuendorf, Kimberly A. (2002): *The Content Analysis Guidebook*. Thousand Oaks u.a.: Sage.
- Rössler, Patrick (2005): *Inhaltsanalyse*. Konstanz: UVK/UTB.
- Scheufele, Bertram (2004): Notwendigkeit, Nutzen und Aufwand von Mehrfach- und Sondercodierungen. In: Wirth, Werner/ Lauf, Edmund (Hrsg.): *Inhaltsanalyse: Perspektiven, Probleme, Potentiale*. Köln: Halem, 82–97.
- Schulz, Winfried (1976): *Die Konstruktion von Realität in den Nachrichtenmedien. Analyse der aktuellen Berichterstattung*. Freiburg/München: Alber.

Medienresonanzanalyse

Eine Einführung in Theorie und Praxis

Raupp, J.; Vogelsang M.A., J.

2009, 210 S. 17 Abb. Mit Online-Extras., Softcover

ISBN: 978-3-531-16000-9