

Lassen sich Ergebnisse aus impliziten Motivtests mit den 100 Worte Motivkategorien vorhersagen: Eine Validierung.

Autor: Daniel Spitzer

Hintergrund: Implizite Motive – also Beweggründe, die menschliches Denken leiten und Handeln steuern – erfahren immer noch Aufmerksamkeit und Interesse (Schultheiss & Pang, 2007). Auch bei der Vorhersage von beruflichen Erfolgsfaktoren wie Arbeitszufriedenheit oder Position innerhalb eines Unternehmens, zeigte sich die prognostische Validität von impliziten Motiven (z. B. Collins, Hanges & Locke, 2004). Es gibt verschiedene Tests, mit denen implizite Motive erfasst werden können. Traditionell werden Versuchspersonen mehrdeutige Bilder vorgelegt, die sie zur Produktion von Geschichten anregen sollen. Es wird angenommen, dass implizite Motive das Denken und Verhalten von Menschen mitbestimmen und daher Niederschlag in den Textproben finden. Diese Textproben werden dann mit Hilfe von Kodierungsregeln von Kodierkräften manuell erfasst. Ein großer Nachteil dieses Verfahren ist augenscheinlich: Der große Aufwand der betrieben werden muss, um die Motive zu erfassen. Dabei stellt bereits Schultheiss (2013) die Frage, ob Motive nicht auch automatisch aus Sprache erfasst werden können. Letztlich sind es auch Texte, die die Grundlage für die traditionelle Kodierung bilden und Motive sollten sich in einzelnen indikativen Worten oder in Kombinationen von Worten abbilden. Ähnlich wie ein Mensch würde auch ein Programm bestimmte Worte oder Wortkombinationen als zugehörig zu einem Motiv kodieren, weil bestimmte Worte wie z. B. „berühmte“ (für n Macht), „weiterentwickelt“ (für n Leistung) und „zusammensein“ (für n Beziehung) dahinterliegende Motive andeuten. Zu klären gilt folglich, ob und wie gut automatische Textanalyse-Programme die Ergebnisse von traditionell kodierten Motiven vorhersagen können. Bereits Schultheiss (2013) untersuchte, wie hoch Kategorien des Linguistic Inquiry and Word Count (LIWC, Version 2001) Motive, die mit dem klassischen Kodiermanual erhoben wurden, korrelieren. Er fand hypothesenkonforme Zusammenhänge, deren Effektstärken sich mehrheitlich im mittleren Bereich bewegten. Seine Ergebnisse sind folgendermaßen zusammengefasst dargestellt:

„Assessment of implicit motives with a word-count approach yields scores that converge with content-coded motive measures, that predict well-documented validation criteria of implicit motive measures, and that respond sensitively to experimental arousal of motivation.“ (Schultheiss, 2013, S.18)

Ziel der vorliegenden Studie ist es, die Validität der 100 Worte Motivkategorien zu untersuchen. Dazu prüfen wir, ob Konvergenz besteht zu klassisch kodierten Motiven. Neben der Prüfung der konvergenten Validität soll untersucht werden, ob sich eine häufig gemachte Beobachtung klassisch kodierter Motive auch mit der 100 Worte Textanalyse replizieren lässt (=Kriteriumsvalidität). In Studien wurde häufig ein Geschlechtereffekt für das Beziehungs-/Anschlussmotiv beobachtet, dergestalt, dass Frauen ein höheres Beziehungsmotiv aufweisen als Männer. Diesen Zusammenhang erwarten wir auch in den von der 100 Worte Textanalyse gemessenen Motiven. Dabei orientieren wir uns an seinem Vorgehen.

Methode: Im Folgenden wird erklärt, welcher Datensatz für die vorliegende Studie ausgewählt wurde, wie sich die 100 Worte Motivkategorien zusammensetzen und wie das als Validierungsmerkmal dienende Außenkriterium entstand.

Ausgangslage: Wie schon erwähnt, untersuchte Schultheiss 2013 bereits, wie gut sich Ergebnisse klassisch kodierter Motive durch Textanalyseprogramme vorhersagen lassen. Die dazu analysierten Daten wurden uns freundlicherweise für unsere Validierungsstudie zur Verfügung gestellt. An dieser Stelle möchten wir uns ausdrücklich bei ihm und seinen Kollegen dafür bedanken. Dieser Datensatz entstand 2009 im Rahmen der Studie „*Activity inhibition: a predictor of lateralized brain function during stress*“. 100 Studenten der Friedrich-Alexander Universität Erlangen nahmen daran teil und absolvierten eine sechs Bilder umfassende „Bilder-Geschichten“ Übung (Picture-Story-Exercise, PSE; McClelland et al., 1989), der in weiten Teilen den Thematischen-Auffassungs-Test widerspiegelt, jedoch im Umfang etwas reduziert ist. Die Teilnehmer wurden instruiert, sich für jedes der sechs mehrdeutigen Bilder eine Geschichte zu überlegen und aufzuschreiben. Anschließend wurden die Geschichten von zwei erfahrenen Kodierkräften mithilfe des Kodiermanuals von Winter (1994) bewertet. Die beiden Kodierungen erreichten eine durchschnittliche Inter-Rater-Reliabilität von 0.80 was als ausreichend hoch betrachtet wurde.

100 Worte Motiv-Korpora: Die 100 Worte Textanalyse enthält Wörterbücher, die implizite Motive erfassen sollen. Diese teilen sich, gemäß der Definition von McClelland, in Wörterbücher auf, die das Machtmotiv, das Leistungsmotiv sowie das Beziehungsmotiv erfassen. Jede dieser drei Motivkategorien unterteilt sich in drei Unterkategorien. Diese bilden für jedes Motiv eine Annäherungs- und eine Vermeidungstendenz ab sowie eine neutrale Kategorie, die zwar Worte enthält, die mit dem zugrundeliegenden Motiv assoziiert sind, jedoch keine Valenz aufweist. Alle Motiv-Wörterbücher umfassen 7438 Worte. Anzumerken ist, dass die 100 Worte Textanalyse kein Stemming verwendet, Worte also nicht auf einen Stamm reduziert und sie daher deutlich mehr Worte enthält als Textanalyseprogramme, die Stemming einsetzen (z. B. LIWC mit 979 Worten in den drei Motivkategorien in der englischen Version von 2015). Die 100 Worte Textanalyse verwendet kein Stemming, da hierdurch Fehler in der Erkennung von Wortbedeutungen entstehen können, worauf Schwartz et al. (2013) hinweisen.

Außenkriterium: Um die konvergente Validität der 100 Worte Motivkategorien zu untersuchen, bedurfte es eines Außenkriteriums, welches implizite Motive auf verlässliche Weise misst. Aufgrund der schwierigen Erfassbarkeit von impliziten Motiven, kommen von vorne herein nur wenige Verfahren in Frage. Neben sprachlich/schriftlichen Verfahren gibt es auch reaktionszeitbasierte Verfahren wie die Motive Superiority Measure (MSM; Eichstaedt, 2003). Wir haben uns letztlich für die Picture-Story-Exercise (PSE) entschieden, weil sie bereits schon einmal als Kriterium zur Validierung von Textanalyseprogrammen eingesetzt wurde und wir unsere Ergebnisse damit vergleichen können.

Ergebnisse: Zunächst wurden interne Konsistenzen für die neun Motiv-Kategorien berechnet sowie die Verteilungsfunktionen analysiert. Es zeigte sich, dass die drei Unterkategorien eine

zu geringe Auftretenswahrscheinlichkeit hatten. Daher fassten wir für alle weiteren Berechnungen die drei Unterkategorien zu jeweils einer Motivkategorie zusammen. So wurden nur noch die Wörterbücher Macht, Leistung und Beziehung weiter untersucht. Die Ergebnisse für die drei aggregierten Wörterbücher sind Tabelle 1 zu entnehmen. Für den Kolmogorow-Smirnow-Test zeigt sich eine Abweichung von der zu erwartenden Normalverteilung. Bei einem Signifikanzniveau von $\alpha=0.05$ und 612 Beobachtungen darf die größte absolute Abweichung den Wert von 0.06 nicht überschreiten. Alle drei Motivkategorien überschreiten diesen Wert (Leistungsmotiv: 0.15; Machtmotiv: 0.14; Beziehungsmotiv: 0.12). Schultheiss beobachtete bei seiner Analyse der LIWC Wörterbücher ebenfalls eine Abweichen von der Normalverteilung. Er empfahl daher eine log-Transformation ($\text{newScore} = \ln(1 + \text{oldScore})$). Bei Kategorien, die eine zu geringe Auftretenswahrscheinlichkeit haben (= Nullausprägung > 20% der Versuchspersonen), wurde die Variable dichotomisiert, also eine Null vergeben, wenn das Merkmal nicht beobachtet wurde und eine 1, wenn das Merkmal beobachtet wurde. Diesem Vorschlag folgen wir und führen die Transformationen entsprechend der beobachteten Verteilung durch. Tabelle 2 ist zu entnehmen, welche Transformationen vorgenommen wurden.

Tabelle 1. *Deskriptive Statistik (M=Mittelwert; SD=Standardabweichung) und interne Konsistenz (=Cronbachs Alpha) für die traditionell kodierten Motive und die 100 Worte Motivkategorien.*

	<i>Alpha</i>	<i>M</i>	<i>SD</i>
<i>PSE [wie bei Schultheiss et al. (2009) berichtet]</i>			
<i>n Macht</i>	0.36	4.55	2.36
<i>n Leistung</i>	0.51	5.14	2.52
<i>n Beziehung</i>	0.48	6.61	2.74
<i>100 Worte</i>			
<i>Macht</i>	0.48	3.07	2.89
<i>Leistung</i>	0.58	3.57	3.46
<i>Beziehung</i>	0.53	4.65	3.92

Die Worterkennungsrate, also das Verhältnis von (mit den 100 Worte Motivkategorien) erkannten Worten im Verhältnis zur Gesamtwortzahl, beträgt 64%. Für die neue Version des LIWC (2015) werden Erkennungsraten von 75% (Texte aus der New-York Times) sowie 92% (Aufgaben zum Expressiven Schreiben) und für die deutsche Version des LIWC (2001) von 63% (Wolf et al., 2008) berichtet.

Um Aufschluss über die konvergente Validität zu erhalten, wurden Korrelationen berechnet. Es zeigen sich signifikante Zusammenhänge zwischen den Ergebnissen den 100 Worte Motivkategorien und den kodierten Motiven. Alle drei Motivkategorien waren in einem mittleren bis hohen Maße positiv mit den kodierten Motiven korreliert (Tabelle 2).

Tabelle 2. Korrelationen zwischen traditionell kodierten Motiven und den 100 Worte Motivkategorien

100 Worte Kategorie	Transformation	n Leistung	n Beziehung	n Macht
Leistung	ln	0.68	-0.37	-0.07
Beziehung	ln	-0.27	0.60	-0.15
Macht	ln	0.23	-0.36	0.41

Anmerkung. Alle Ergebnisse sind auf dem 0.001 Niveau signifikant

Bei der Prüfung der Unabhängigkeit der 100 Worte Motivkategorien in der Vorhersage der traditionell kodierten Motiven, fiel eine nicht zu verachtende Überlappung auf. Diese reichten von 0.07 (für 100 Worte Leistung auf n Macht) bis -0.37 (für 100 Worte Leistung auf n Beziehung). Dieser Befund wird im Teil „Konklusion“ näher beleuchtet.

Neben der Prüfung der konvergenten Validität wurde auch der Zusammenhang des Geschlechts mit dem Beziehungsmotiv untersucht um so auch Hinweise für die Kriteriumsvalidität zu erhalten. Erwartet wurde, dass ein positiver Zusammenhang besteht. Dieser Zusammenhang konnte beobachtet werden. Das Geschlecht korrelierte positiv mit dem 100 Worte Beziehungsmotiv ($t(612)=0.12$, $p<0.005$), jedoch nicht mit den anderen Motivkategorien. Dieser Zusammenhang zeigt sich auch etwas schwächer mit dem klassisch kodierten Beziehungsmotiv ($t(612)=0.11$, $p<0.005$).

Konklusion: Zielsetzung dieser Studie war es zu zeigen, ob die automatisierte Erfassung von Worten und Wortkombinationen mit der 100 Worte Textanalyse Aufschluss über klassisch kodierte Motive geben kann. Dazu wurden die im Schultheiss-Datensatz zur Verfügung gestellten Texte mit den 100 Worte Motivkategorien untersucht und anschließend die konvergente und die Kriteriumsvalidität berechnet. Es wurde vermutet, dass sich signifikant positive Korrelationen zwischen den 100 Worte Motivkategorien und den kodierten Motiven zeigen und das Geschlecht mit dem Beziehungsmotiv ebenfalls positiv zusammenhängt. Die Koeffizienten in der Untersuchung der konvergenten Validität bewegen sich dabei in einem mittlere bis hohen und im niedrigen Bereich für die Kriteriumsvalidität. Dieser war etwas höher als der Zusammenhang zwischen dem Geschlecht und dem klassisch kodierten Beziehungsmotiv. Anzumerken ist hier, dass Zusammenhänge mit einem Außenkriterium (wie dem Geschlecht) oftmals kleiner sind als Zusammenhänge zwischen verschiedenen Tests.

Die in der vorliegenden Studie beobachteten konvergenten Validitätsangaben sind höher als die bei Pennebaker und King (1999) und bei Schultheiss (2013) berichteten.

Um die Güte der 100 Worte Motivkategorien zu bestimmen, scheint auch ein Vergleich mit Ergebnissen von Winter (1991) wo er Konvergenzen von zwei Versionen seines Motiv-Kodier-Manuals untersuchte, sinnvoll. Der Zusammenhang zwischen zwei Versionen reicht von 0.45 bis 0.72. Die in dieser Studie gefundenen Effektstärken bewegen sich in einem ähnlichen Bereich. Die Ergebnisse von Winter zeigen, dass auch unter traditionellen Kodiermanualen keine perfekte Konvergenz besteht.

Gemeinhin wird Unabhängigkeit der Motive voneinander gefordert. Dies war in der vorliegenden Untersuchung, als auch bei Schultheiss (zwischen 0,06 bis 0,49) nicht gegeben. In der vorliegenden Studie waren die 100 Worte Motive Leistung und Macht negativ assoziiert mit n Beziehung. In einem schwächeren Ausmaß das 100 Worte Motiv Beziehung mit n Leistung und n Macht. Zu beachten gilt, dass auch die traditionell kodierten Motive im untersuchten Datensatz Varianz teilen. So kovariieren n Leistung und n Beziehung zu -0,34 und n Macht und n Beziehung zu -0,17. Es scheint, als stünden sich die Motive Leistung und Beziehung gegenüber. In einem geringeren Ausmaß gilt das auch für die Beziehung zwischen den Motiven Macht und Beziehung.

Die Erkennungsrate der 100 Worte Textanalyse liegt bei 64% und ist damit vergleichbar mit den Werten von LIWC.

Insgesamt kann festgehalten werden, dass die 100 Worte Motivkategorien gut geeignet sind, um klassisch kodierte Motive vorherzusagen. Schwierigkeiten mit der Reliabilität sowie der Unabhängigkeit der Kategorien bestehen nicht nur bei der 100 Worte Textanalyse, sondern auch bei LIWC und sollten in der weiteren Entwicklung beachtet und möglichst verbessert werden.

Literatur

- Collins, C.J., Hanges, P.J. & Locke, E.A. (2004). The relationship of achievement motivation to entrepreneurial behaviour: A meta-analysis. *Human Performance*, 17, 95-117.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *J. Pers. Soc. Psychol.* 77, 1296–1312.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schultheiss, O.C. & Pang, J.S. (2007). Measuring implicit motives. In R.W. Robins, R.C. Fraley & R.F. Krueger (Hrsg.), *Handbook of research methods in personality psychology* (S. 322-344). New York: Guilford Press.
- Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In J.M. Stewart, J.M. Healy, Jr., & D.J. Ozer (Eds.), *Perspectives in personality, Vol. 3: Teil B: Approaches to understanding lives*. (S. 59-89). Bristol: Jessica Kingsley Publishers.
- Winter, D.G. (1994). *Manual for scoring motive imagery in running text (Version 4.2)*. University of Michigan.
- Wolf, M., Horn, A., Mehl, M., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. *Diagnostica* 2, 85–98.