

IST DIE BILDUNG EINES EMPTHIESCORES IN DER DEUTSCHEN FASSUNG DES IRI SINNVOLL?

Dr. Christoph Paulus
Universität des Saarlandes
FR Bildungswissenschaft
© 2012

Zusammenfassung

Bei der Verwendung des SPF(IRI) zur Messung von Empathie stellt sich immer wieder die Frage, ob man aus testökonomischen Gründen einen Gesamtscore der im Test gemessenen Skalen (empathic concern [EC], perspective taking [PT], personal distress [PD], fantasy scale [FS]) vornehmen kann. Im englischen Original des Fragebogens (IRI) von Davis (1983) ist dies nicht vorgesehen. Im vorliegenden Artikel konnte jedoch anhand der Analysen von ROC-Kurven verschiedener Scorebildungen gezeigt werden, dass eine solche durchaus sinnvoll durchgeführt werden kann als $E = EC + PT + FS$

Summary

By the use of the SPF (IRI) to the measurement of empathy the question positions itself over and over again whether one can carry out a general score of the scales measured in the test (empathic concern [EC], perspective taking [PT], personal distress [PD], fantasy scale [FS]). In the English original version of the questionnaire (IRI) by Davis (1983) this is not planned. Nevertheless, in the present article could be shown with the help of the analyses of ROC curves of different scoring methods that this is possible to score $E = EC + PT + FS$

Einleitung

Seit den 1980er Jahren teilt man Empathie in zwei Bereiche „kognitive Empathie, also die Fähigkeit, mental in die Haut eines anderen zu schlüpfen und affektive Empathie als die emotionale Reaktion auf einen beobachteten Gemütszustand. Letzteres entspricht dem, was man unter Mitgefühl versteht.“ (Kraft & Dziobek, 2006).

Der vom Sozialpsychologen Mark H. Davis (1983) entwickelten interpersonellen Reaktivitätsindex (IRI) erfasst beide Bereiche. Er unterteilte bei der Konstruktion des IRI, ausgehend von der Überlegung, dass sowohl affektive als auch kognitive Komponenten bei empathischer Reaktion vorhanden sein können, Empathie in vier miteinander in Beziehung stehende Subgruppen: *perspective taking* (PT), *fantasy* (FS), *empathic concern* (EC) und *personal distress* (PD). Somit wird auch die von Batson (1989, 1991) vorgeschlagene Unterscheidung zwischen egoistischer (*personal distress*) und altruistischer (*empathic concern*) Emotion bzw. Motivation berücksichtigt.

PT misst dabei die Fähigkeit, spontan eine Sache aus der psychologischen Perspektive eines anderen sehen zu können; die *fantasy scale* (FS) erfasst die Tendenz der Vp, sich in die Rolle und Handlungsweise von Figuren in Romanen oder Filmen zu versetzen. Die restlichen beiden Subskalen stellen Operationalisierungen der typischen emotionalen Reaktionen eines Beobachters dar: Die EC-Scala dient zur Messung fremd-orientierter Gefühle wie Mitleid oder Sorge um Personen in Not, die PD-Scala dagegen soll eigen fokussierte Gefühle wie Unruhe oder Unwohlsein in engen interpersonalen Beziehungen messen.

Die wenigen deutschen Übersetzungen weisen allerdings ähnliche testtheoretischen Schwächen auf wie das Original (vgl. dazu z.B. Enzmann, 1996; Beven, 2004; Lauterbach & Hosser, 2007). Nach einer mehrfachen Überarbeitung aufgrund faktorenanalytischer Ergebnisse und marginalen Umformulierungen in der Übersetzung konnte Paulus (2006) eine deutsche Version der IRI erstellen, die gute Kennwerte in den Bereichen der Reliabilität, der faktoriellen Validität und der Itemtrennschärfen aufweist. Die interne und externe Validität konnte durch Heranziehen von Vergleichswerten aus existierenden Empathiefragebögen sowie einer Selbsteinschätzung der Pbn gesichert werden.

Korrelationen

		FS	EC	PT	PD
FS	Korrelation nach Pearson	1	,457**	,322**	,207**
	Signifikanz (2-seitig)		,000	,000	,000
	N	4410	4410	4410	4410
EC	Korrelation nach Pearson	,457**	1	,478**	,224**
	Signifikanz (2-seitig)	,000		,000	,000
	N	4410	4410	4410	4410
PT	Korrelation nach Pearson	,322**	,478**	1	-,028
	Signifikanz (2-seitig)	,000	,000		,068
	N	4410	4410	4410	4410
PD	Korrelation nach Pearson	,207**	,224**	-,028	1
	Signifikanz (2-seitig)	,000	,000	,068	
	N	4410	4410	4410	4410

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Tabelle 1: Skaleninterkorrelation des SPF

Da die 4 Skalen nicht unabhängig voneinander sind, stellt sich die Frage, ob eine Summenscorebildung theoretisch möglich sei. Bisher wird von diesem Vorgehen abgeraten, weil die Faktoren als weitgehend unabhängig voneinander angesehen wurden (vgl. z.B. Rogers, 2007) bzw. es wurde nach unterschiedlichen Methoden aus testökonomischen Gründen ein Summenscore gebildet (z.B. Cliffordson, 2001, 2002; Hülsmann, 2008).

Die Tabelle der Inter-Skalen-Korrelationen (s. Tabelle 1) zeigt dagegen deutlich, dass von einer weitgehenden Unabhängigkeit der Faktoren in der deutschen Version nicht ausgegangen werden kann. Die Korrelationen sind zwar unterschiedlich stark, aber dennoch nachweisbar.

1.1 Mögliche Empathiescores

Cliffordson (2001) formulierte einen generellen Empathie-Score durch Summierung der Skalen EC+PT+FS, wobei sie die Validität der PD-Skala generell in Frage stellte. Dabei werden die drei am stärksten miteinander korrelierenden Skalen additiv verbunden.

Paulus entwickelte (1997) in einem ersten Ansatz den Empathiescore $E=EC+PT+FS-PD$ mit dem Gedanken, dass *personal distress* insbesondere in Hilfesituationen negative Auswirkungen auf empathiegestützten Altruismus besäße und insofern einen handlungsrelevanten Hemmungsmechanismus darstellen würde.

Davis (1994) selbst drückt seine Zweifel dagegen eher an der FS aus, indem er schreibt: „*The thrust of the FS items (...) suggests that the scale should fall under the heading of role taking. However the fact*

that virtually all of these items deal with imagining oneself in the place of fictitious characters (...) complicates this interpretation" (S. 58), sodass ein weiterer Score durch $E=EC+PT+PD$ möglich wäre.

Im Folgenden wollen wir deshalb die Aussagekraft der drei hier beschriebenen Scores näher betrachten:

- $E_{CLIFF} = EC + PT + FS$ (nach Cliffordson, 2001)
- $E_{PAU} = EC + PT + FS - PD$
- $E_{DAV} = EC + PT + PD$

Methode

1.2 Stichprobe

Den Analysen lag eine Stichprobe von ursprünglich 2.657 Vpn zugrunde, die den SPF teilweise online, teilweise in paper-pencil-Form beantwortet haben. Ergänzend zu den 16 Items des Fragebogens wurden noch 5 Items einer Selbsteinschätzungsskala („self“) hinzugefügt, sowie Angaben zu Alter, Beruf und Schulbildung.

Das mittlere Alter der Pbn war 26,3 Jahre ($s = 11,43$) und reichte von 10 bis 80 Jahren; 28,7% ($n=766$) waren männlich und 71,3% weiblich. Diese Ungleichgewichtung ist allerdings nicht sehr relevant, wenn man sich die Mittelwertsunterschiede ansieht (Tabelle 2). Obwohl sich wegen der Stichprobengröße mit Ausnahme der Skala PT alle anderen Mittelwerte signifikant unterscheiden, sind die Effektstärken so gering, dass eine künstliche Verkleinerung der weiblichen Stichprobe nicht notwendig erscheint.

	sex	Mittelwert	s	N	t / df	Effektstärke
self	männlich	18,13	3,370	766		
	weiblich	18,81	3,008	1901	-4,82 ($p<.00$) df = 1282,01	0,01
FS	männlich	13,66	3,314	766		
	weiblich	15,14	3,070	1901	-10,62 ($p<.00$) df=1321,92	0,04
EC	männlich	13,98	3,067	766		
	weiblich	15,00	2,808	1901	-7,95 ($p<.00$) df=2665	0,02
PT	männlich	14,51	3,132	766		
	weiblich	14,32	3,100	1901	1,38 (n.s.)	0,00

	sex	Mittelwert	s	N	t / df	Effektstärke
PD	männlich	10,80	3,286	766		
	weiblich	12,06	3,233	1901	-9,01 (p<.00) df=1393,10	0,03

Tabelle 2: Mittelwerte und Effektstärken der Skalen

1.3 Materialien

Es wurde die Fassung V5.7 der Saarbrücker Persönlichkeitsskala SPF¹ (Paulus, 2006), die eine deutsche Übersetzung des Interpersonal Reactivity Index nach Davis (1980) darstellt, benutzt. Diese misst mit je 4 Items mit sehr guten Gütekennwerten² die Skalen empathic concern, perspective taking, fantasy und personal distress. Zusätzlich wurden 5 Items zur Selbstbeschreibung mit erhoben, die eine subjektive Einschätzung der eigenen Empathie beschreiben sollen. Als Antwortvorgabe diente eine 5-stufige Antwortskala (wie häufig trifft die geschilderte Situation auf Sie zu?) von „nie“ (1) bis „immer“ (5).

- Ich würde mich selbst als einen einfühlsamen Menschen bezeichnen.
- Ich kann anderen Menschen geduldig zuhören.
- Ich habe Verständnis für Eitelkeiten oder bestimmte Verhaltensweisen wie Profiergehebe, Rechthaberei, Überheblichkeit o.ä. bei Freunden oder Bekannten.
- Ich verstehe die Gefühle anderer und achte auf deren Bedürfnisse.
- Die Körpersprache anderer Menschen sagt mehr aus als ihre Worte.

Tabelle 3: Items zur Selbstbeschreibung

1.4 Analyse

Bei der Diagnostik von Persönlichkeitseigenschaften kann man im Grunde genommen zwei Fehler begehen: Man kann geeignete Personen als ungeeignet klassifizieren (falsch negativ) und ungeeignete Personen als geeignet beschreiben (falsch positiv). Die Güte eines diagnostischen Kennwertes misst sich u.a. nun daran, wie hoch seine Sensitivität und seine Spezifität ist.

¹ <http://bildungswissenschaften.uni-saarland.de/personal/paulus/empathy/SPF.html>

² <http://bildungswissenschaften.uni-saarland.de/personal/paulus/empathy/TTSPF%28IRI%29.htm>

Als **Sensitivität** bezeichnet man dabei den Anteil der richtig-positiven Diagnosen an der Gesamtzahl der Probanden, sie beschreibt also die Eignung des Tests, positive Personen möglichst genau herauszufiltern. Die **Spezifität** eines Tests beschreibt, wie gut der Anteil der richtig-negativen Pbn erkannt werden kann. Beide Maße dienen also der Beschreibung, wie gut die Diagnosefähigkeit eines Tests ist. Problematisch ist dabei allerdings immer der sog. cut-off-Punkt, d.h. derjenige Wert, ab dem ein Testwert diese Bedingungen in hinreichendem Maße erfüllen kann. Um diesen mit einer bestimmten Irrtumswahrscheinlichkeit zu bestimmen, kann man auf die ROC-Methode zurückgreifen, die ursprünglich aus der Elektrotechnik stammt.

„Die Receiver Operating Characteristic (ROC) – Kurve bzw. Grenzwertoptimierungskurve ist eine Methode zur Bewertung und Optimierung von Analyse-Strategien. Die ROC-Kurve stellt visuell die Abhängigkeit der Effizienz mit der Fehlerrate für verschiedene Parameterwerte dar. Für jeden möglichen Parameterwert (z. B. Sendegeschwindigkeit, Frequenz, ...) ermittelt man die resultierenden relativen Häufigkeitsverteilungen in Form von Sensitivität und Falsch-Positiv-Rate. In einem Diagramm trägt man Sensitivität als Ordinate und Falsch-Positiv-Rate als Abszisse ein. Der Parameterwert selbst taucht dabei nicht auf, kann aber als Beschriftung der Punkte verwendet werden. Es resultiert typischerweise eine gekrümmte, aufsteigende Kurve. Zu der ROC-Kurve berechnet man die Fläche unterhalb der Kurve („Area under curve“, ROC AUC). Dieser Wert kann zwischen 0 und 1 liegen, wobei aber 0,5 der schlechteste Wert ist. Wie zuvor beschrieben ist eine ROC-Kurve nahe der Diagonalen das zu erwartende Ergebnis eines Zufallsprozesses, der eine Fläche von 0,5 hat. Die zuvor als optimal beschriebene Kurve hat eine Fläche zwischen 0,5 und 1.“ (wikipedia).

Dabei zeigt die Abweichung nach oben von der Diagonalen (also eine Zunahme der Fläche unter der Kurve) an, wie gut die Diagnoseeigenschaften des verwendeten Testmaßes sind.

Für die drei von uns beschriebenen Empathie Maße E_{CLIFF} , E_{PAU} und E_{DAV} bestimmen wir zunächst das Außenkriterium. Dazu benutzen wir den Score der Selbsteinschätzung als Ausgangsmaß. Die Eignung von Selbsteinschätzungen als Kriterium für die Empathiemessung wird u.a. bei Cliffordson (2001) beschrieben.

In unserer Stichprobe ist der Mittelwert der Selbsteinschätzungsskala bei 18,64 ($s = 3,09$); an diesem Wert dichotomisieren wir die Selbsteinschätzung an der Grenze $MW+s$ ($18,64+3,09$) und bezeichnen damit die oberen knapp 15% der Stichprobe als empathisch (=1), alle anderen auftretenden Einschätzungen werden als „wenig oder kaum empathisch (=0) klassifiziert. Damit setzen wir eine recht harte Kriteriumsgrenze an, vermeiden damit aber vorschnelle positive Eingruppierungen.

Ergebnisse

Betrachtet man nun die ROC-Kurven der drei beschriebenen Empathie­maße, so zeigt sich deutlich, dass die Fläche unter der Kurve bei E_{CLIFF} am größten wird, d.h. dieses Maß zur Bestimmung ist von den drei zur Auswahl stehenden Möglichkeiten aus diagnostischer Sicht am besten geeignet.

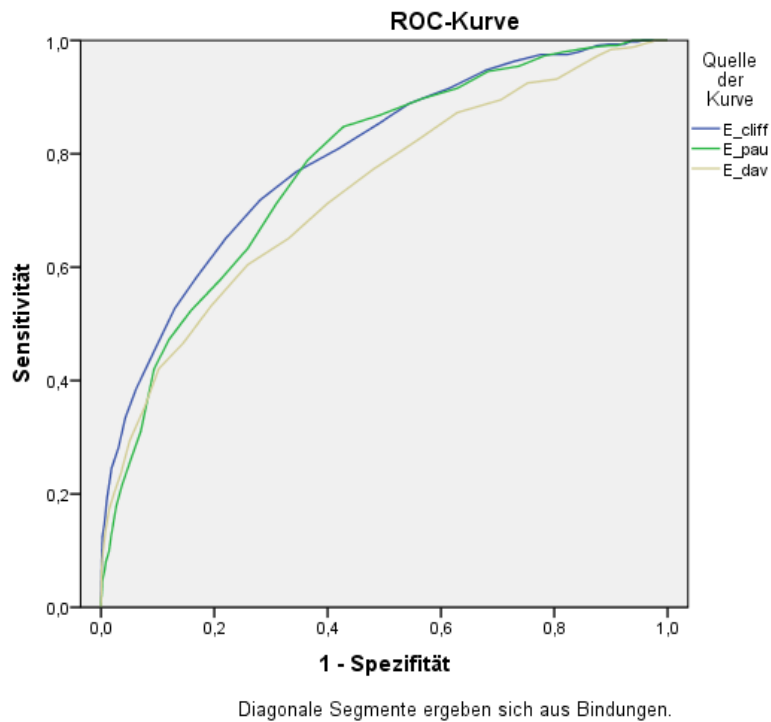


Abbildung 1: ROC-Kurven möglicher Empathie­maße

Fläche unter der Kurve

Variable(n) für Testergebnis	Fläche	Standardfehler	Asymptotische Signifikanz	Asymptotisches 95% Konfidenzintervall	
				Untergrenze	Obergrenze
E_cliff	,793	,012	,000	,770	,816
E_pau	,776	,012	,000	,753	,799
E_dav	,732	,014	,000	,705	,759

Tabelle 4: Flächenbestimmung möglicher Empathie­maße

Betrachten wir als E_{CLIFF} in seinen möglichen Ausprägungen, so ergibt sich folgende Liste (Auszug):

Koordinaten der Kurve
Variable(n) für Testergebnis: E_{cliff}

Positiv, wenn größer oder gleich	Sensitivität	1 - Spezifität
43,5000	,855	,492
44,5000	,809	,419
45,5000	,768	,345
46,5000	,718	,281

Tabelle 5: Auszug aus den Koordinaten der Kurve zu E_{CLIFF}

Mit einer Irrtumswahrscheinlichkeit von 5% liegt der optimale cut-off-point bei ca. 45 Punkten (entspricht einer Sensitivitätsgrenze von 0,77; s. Tabelle 4). Damit können alle Pbn mit einem Summenwert $E_{\text{CLIFF}} \geq 45$ Punkten mit einer Irrtumswahrscheinlichkeit von 5% als empathisch eingestuft werden.

Diskussion und Zusammenfassung

Bei der Verwendung des SPF(IRI) zur Messung von Empathie stellt sich immer wieder die Frage, ob man aus testökonomischen Gründen einen Gesamtscore der im Test gemessenen Skalen (empathic concern, perspective taking, personal distress, fantasy scale) vornehmen kann. Im englischen Original des Fragebogens (IRI) von Davis (1983) ist dies nicht vorgesehen. Im vorliegenden Artikel konnte jedoch anhand der Analysen von ROC-Kurven verschiedener Scorebildungen gezeigt werden, dass eine solche durchaus sinnvoll durchgeführt werden kann als $E = EC + PT + FS$.

Die Einbeziehung Skala PD verringert dabei die Sensitivität, unabhängig davon, ob diese additiv oder subtraktiv (wie in E_{PAU}) berechnet wird.

Fläche unter der Kurve

Variable(n) für Testergebnis	Fläche	Standardfehler	Asymptotische Signifikanz	Asymptotisches 95% Konfidenzintervall	
				Untergrenze	Obergrenze
E _{CLIFF}	,793	,012	,000	,770	,816
PD_plus	,746	,013	,000	,720	,772
PD_minus	,776	,012	,000	,753	,799

Tabelle 6: Flächen unter der Kurve unter Einbeziehung der Skala PD

Personal distress als negativer Einfluss auf Empathie erscheint nach den Daten der ROC-Kurve noch sinnvoller, obwohl dieser Score nicht die Güte von E_{CLIFF} erreicht. Erklärt werden könnte dies dadurch, dass PD, so wie ursprünglich von Davis (1994) beschrieben, „the tendency to experience distress and discomfort in response to extreme distress in others (S. 57)“, inhaltlich sehr eng sowohl mit EC als auch mit PT verbunden sein könnte, obwohl dies die o.g. Korrelationen (vgl. **Fehler! Verweisquelle konnte nicht gefunden werden.**) nicht belegen können. Damit wird erneut die Validitätsproblematik der PD-Skala angesprochen, auf die Cliffordson (2001) schon hingewiesen hat.

Überhaupt scheinen die Ergebnisse zu PD recht uneinheitlich zu sein: So konnten z.B. Coke et al. (1978) oder Toi & Batson (1982) zeigen, dass EC gleichzeitig den Prozess des PD anregt und dass Testpersonen beide affektiven Zustände gleichzeitig empfanden. Während bereits recht früh Stotland & Dunn (1963) oder Krebs (1975) diese Ähnlichkeit zwischen EC und PD bei induzierten Angstsituationen zeigen konnten, fanden Marks et al. (1982) keinerlei Effekte in dieser Hinsicht. In neueren Studien wies Houston (1990) nach, dass eine experimentelle Aktivierung von EC nicht automatisch auch PD anregte. Das Konzept PD mit seiner Beziehung zu EC bleibt also unklar und bedarf weiterer Untersuchungen.

Eine weitere offene Frage bleibt, in wie weit das von uns gebildete Außenkriterium über Selbsteinschätzungssitems und die damit verbundene Berechnung von Sensitivität und Spezifität eines Empathiescores valide ist. Mögliche weitere Ansätze in diesem Feld wären eine Änderung der Selbsteinschätzungsscores gemäß Cliffordson (2001) bzw. ein externes Kriterium durch den Einsatz weiterer Empathietests. Ebenso ist es diskussionswürdig, ob die Grenzsetzung des Außenkriteriums bei MW+s geeignet erscheint oder nicht vielleicht sogar als überzogen kritisch anzusehen wäre.

Literatur

- Batson, C.D., & Oleson, K.C. (1991): Current status of the empathy-altruism hypothesis. In: M.S. Clarke (Ed.): *Prosocial behavior*. London: Sage Publ., 62-85.
- Batson, C.D., Oleson, K.C., Weeks, J.L., Healy, S.P., Reeves, P.J., Jennings, P, & Brown, T. (1989): Religious prosocial motivation: Is it altruistic or egoistic? *Journal of Personality and Social Psychology*, 57, 873-884.
- Cliffordson, C. (2001). Parents' judgment and students' self-judgments of empathy. *European Journal of Psychology Assessment*, 17, 36-47.
- Cliffordson, C. (2002). Interviewer agreement in the judge of empathy in selection interviews. *International Journal of Selection and Assessment*, 10, 198-205.
- Coke, J.S., Batson, C.D. & McDavis, K. (1978). Empathic mediation of helping. A two-stage model. *JPSP*, 36, 752-766.
- Davis, M. (1983): Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- Davis, M. (1994). *Empathy. A Social Psychological Approach*. Madison, WI: Brown & Benchmark's.
- Houston, D. A. (1990). Empathy and the self: Cognitive and emotional influences on the evaluation of negative affects in others. *JPSP*, 59, 859-868.
- Hülsmann, A. (2008). Empathiefähigkeit schizophrener Patienten. URL: [thes.univie.ac.at/2505/1/2008-11-06_0304338.pdf](https://theses.univie.ac.at/2505/1/2008-11-06_0304338.pdf)
- Krebs, D.L. (1975). Empathy and altruism. *JPSP*, 32, 1134-1146.
- Kraft, U. & Dziobek, I. (2006). *Das Mitleid der Teilnahmslosen*. http://www.handelsblatt.com/news/-printpage.aspx?_p=301118&_t=ftprint&_b=1129076.
- Marx, E. L., Penner, L.A., & Stone, A.V.W. (1982). Helping as a function of empathic response and sociopathy. *Journal of Research in Personality*, 16, 1-20.
- Paulus, C. (1997). Empathie, Kompetenz und Altruismus. <http://www.uni-saarland.de/fak5/ezw/personal/paulus/empathie.htm>
- Paulus, C. (2006). Der Saarbrücker Persönlichkeitsfragebogen (SPF-IRI) zur Messung von Empathie. <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/>
- Stotland, E. & Dunn, R.E. (1963). Empathy, self-esteem, and birth-order. *Journal of Abnormal and Social Psychology*, 66, 532-540.

Toi, M. & Batson, C.D. (1982). More evidence that empathy is a source of altruistic motivation. *JPSP*, 43, 281-292.