

Aus dem Forschungsprojekt des Arbeitsbereichs „Psychoanalytische Pädagogik“  
des Instituts für Bildungswissenschaft der Universität Wien:

## THEOPRAX

Die Bedeutung von Theorie für psychosoziale Praxis aus der Sicht von Akteur\*innen:  
Eine Untersuchung der mentalen Repräsentation praxisleitender Theorien

Projektleitung: Univ.-Prof. Dr. Wilfried Datler // Projektkoordination: Univ.-Ass. Bernadette Strobl, MA  
Kontakt: bernadette.strobl@univie.ac.at; sonja.scheidl@univie.ac.at

-----

**Tamara Katschnig<sup>1</sup> und Corinna Geppert<sup>2</sup> (2019):**

**Die Bestimmung der Interrater-Reliabilität in einem TheoPrax-Teilprojekt.  
Ein Beitrag zur Auswertung von 5 WIRTH-Interviews in Hinblick auf  
5 Dimensionen samt Subdimensionen.**

Institut für Bildungswissenschaft der Universität Wien und  
Institut Fortbildung für PädagogInnen der KPH Wien/Krems

Mit Ergänzungen vom 8.2.2021.

-----

*Im Forschungsprojekt TheoPrax wird unter anderem untersucht, wie psychosozial Tätige in einem eigens konzipierten Interview namens WIRTH der Bitte bzw. Aufforderung nachkommen, an einem konkreten Praxisbeispiel darzulegen, in welcher Weise sie sich in ihrer Arbeit an Theorien oder Konzepten orientieren. Die Antworten werden nach verschiedenen Gesichtspunkten geratet. Erste nähere Informationen zum Projekt TheoPrax können bei Strobl (2017, 2019) sowie Strobl/Datler (2020, 2021) nachgelesen werden.*

*Die Auswertung der Antworten, welche in Interview gegeben werden, erfolgt jeweils in Teilprojekten durch Rater\*innen. Nachdem bereits 2017 im Vorfeld die Interrater-Reliabilität der Rater\*innen bestimmt worden war, die Zusammensetzung des Teams der Rater\*innen sich aber veränderte, wurde neuerlich die Interrater-Reliabilität bestimmt. Darüber informiert der vorliegende Text.*

---

<sup>1</sup> HS-Prof. Mag. Dr. Priv.-Doz. Tamara Katschnig, Institut Fortbildung für PädagogInnen der KPH Wien/Krems sowie Institut für Bildungswissenschaft der Universität Wien

<sup>2</sup> MMag. Dr. Corinna Geppert, Institut für Bildungswissenschaft der Universität Wien

Als Interrater-Reliabilität wird die Übereinstimmung zwischen zwei oder mehreren Beurteiler\*innen zu gleichen Fällen bezeichnet, wodurch angegeben wird, in welchem Ausmaß die Ergebnisse von Beurteilungen (Ratings) unabhängig davon anzusehen sind, welche Beurteiler\*innen welche Beurteilung abgegeben haben. Ziel ist es, durch Proberatings zu einem „einheitlichen“ Ratingverfahren zu gelangen, um unterschiedliche Beurteilungen innerhalb einer Gruppe von Beurteiler\*innen immer unwahrscheinlicher werden zu lassen. Kommen zwei (oder mehr) Beurteiler\*innen unabhängig voneinander zur gleichen Einschätzung, kann von hoher Interrater-Reliabilität gesprochen werden. Diese liefert Aussagen hinsichtlich der Genauigkeit von Urteilen und (implizit) auch ihrer Objektivität. Die Erfassung und Untersuchung der Interrater-Reliabilität setzt daher voraus, dass mindestens zwei Rater\*innen Urteile zu denselben Objekten abgegeben haben.

Im Rahmen des Projekts „TheoPrax“ gab es einen Termin (20.12.2018), an dem 4 Rater\*innen (drei Studierende und eine Absolventin der Bildungswissenschaft<sup>3</sup>) jeweils 5 Interviews (im Folgenden „Fälle“ genannt) in Hinblick auf 5 Dimensionen mit insgesamt 19 Subdimensionen rateten:

- Dimension 1 umfasst 3 Subdimensionen
- Dimension 2 umfasst 4 Subdimensionen
- Dimension 3 umfasst 4 Subdimensionen
- Dimension 4 umfasst 4 Subdimensionen
- Dimension 5 umfasst 4 Subdimensionen

Die Rater\*innen hatten sich im Rahmen des Ratings entlang einer jeden Dimension für jeweils eine Subdimension durch die Vergabe definierter Signaturen zu entscheiden. (Nähere Informationen über die Dimensionen, die Subdimensionen und Signaturen sind in einer ersten Fassung bei Strobl 2017 nachzulesen. Erläuterungen zu den Dimensionen, nach denen geratet wird, sind bei Strobl/Datler 2021 zu finden.)

Tabelle 1 zeigt auf der nachfolgenden Seite die Werte des Cronbachs Alpha der Rater\*innen in den 5 Dimensionen sowie insgesamt, wobei Werte ab 0,8 als „gut“ gelten.

---

<sup>3</sup> An dieser Stelle sei den Rater\*innen Stefanie Maria Pichler, Manuel Prem, Sara Wieneke und Bernadette Strobl herzlich gedankt.

Fälle 1-5	Cronbachs alpha					Gesamt
	Dimension1	Dimension2	Dimension3	Dimension4	Dimension 5	
4 Rater*innen	1	.712	.990	.675	.786	.869

Tabelle 1: Cronbachs Alpha Werte

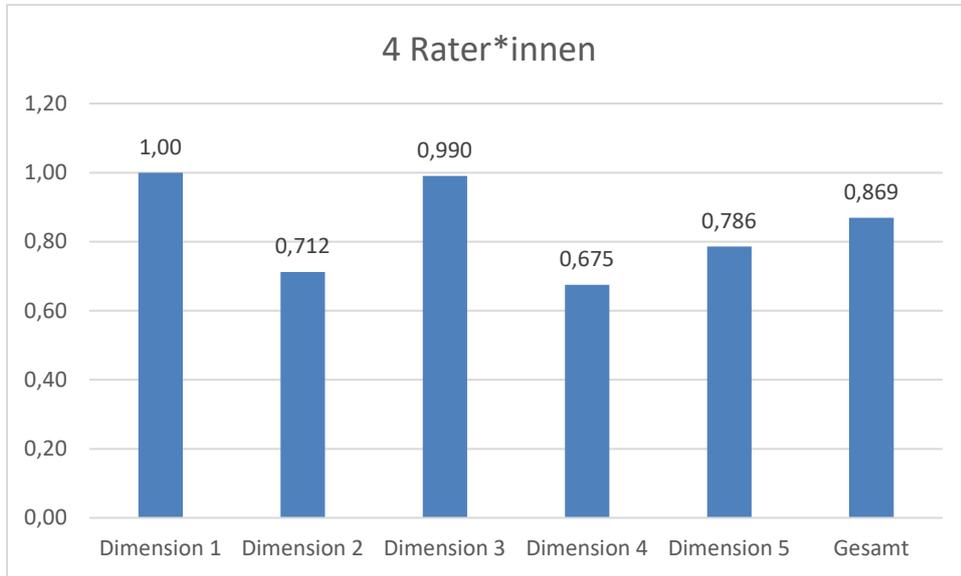


Abbildung 1: Cronbachs Alphawerte der 4 Rater\*innen (Fälle 1-5) und Gesamt

Abbildung 1 und Tabelle 1 zeigen die Reliabilität der Ratings der Fälle 1 bis 5, bezogen auf 4 Rater\*innen. Auffallend ist, dass es in der Dimension 1 und 3 kaum Unterschiede zwischen den Urteilen der 4 Rater\*innen gibt, sich diese jedoch in den Dimensionen 2, 4 und 5 doch unterscheiden. Insbesondere Dimension 4 scheint Probleme beim Raten zu verursachen. Der Gesamtwert ist mit .869 in einem guten Bereich.

Konkrete Analysen zu den einzelnen Dimensionen: Bei Dimension 1 gibt es 100%ige Übereinstimmung bei allen Rater\*innen. In den folgenden Tabellen 2 bis 5 sind die Item-Skala-Statistiken dargestellt. Die letzte Spalte gibt jeweils an, wie sehr sich die Reliabilität der einzelnen Ratings in den Dimensionen erhöhen würde, wenn ein\*e Rater\*in nicht bewerten würde.

Item-Skala-Statistiken				
	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Cronbachs Alpha, wenn Item weggelassen
rater1_2	8,80	12,700	,500	,661
rater2_2	8,80	6,700	,987	,246
rater3_2	8,80	12,700	,500	,661
rater4_2	8,40	11,800	,228	,852

Tabella 2: Reliabilität Dimension 2

Bei Dimension 2 fällt auf, dass Rater\*in 4 sich von den Rater\*innen 1, 2 und 3 unterscheidet.

Item-Skala-Statistiken				
	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Cronbachs Alpha, wenn Item weggelassen
rater1_3	7,80	19,700	,988	,982
rater2_3	7,80	19,700	,988	,982
rater3_3	8,00	21,500	,964	,991
rater4_3	7,60	18,800	,968	,989

Tabella 3: Reliabilität Dimension 3

Bei Dimension 3 fällt auf, dass aller Rater\*innen sehr nahe aneinander/ähnlich geratet haben.

Item-Skala-Statistiken				
	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Cronbachs Alpha, wenn Item weggelassen
rater1_4	5,40	12,300	,233	,732
rater2_4	6,40	10,300	,386	,655
rater3_4	7,00	10,000	,884	,450
rater4_4	6,40	6,800	,564	,551

Tabella 4: Reliabilität Dimension 4

Bei Dimension 4 fällt auf, dass Rater\*in 1 sich von den Rater\*innen 2, 3 und 4 unterscheidet.

Item-Skala-Statistiken				
	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Cronbachs Alpha, wenn Item weggelassen
rater1_5	5,00	18,500	,212	,876
rater2_5	5,00	10,000	,866	,570
rater3_5	5,40	15,300	,800	,706
rater4_5	5,00	8,500	,753	,671

Tabella 5: Reliabilität Dimension 5

Bei Dimension 5 fällt auf, dass Rater\*in 1 sich von den Rater\*innen 2, 3 und 4 unterscheidet. Daraufhin wurde ermittelt, in welchem Ausmaß es zwischen Rater\*innen (Paare) Abweichungen in der Bewertung gibt.

Abbildung 2 zeigt den Gesamtscore der Abweichungen der Rater\*innen jeweils paarweise, wobei ersichtlich ist, dass Rater\*in 1 und Rater\*in 3 am wenigsten voneinander abweichen (8 mal von möglichen 19 Abweichungen) und Rater\*in 3 und Rater\*in 4 am meisten (13 von 19).

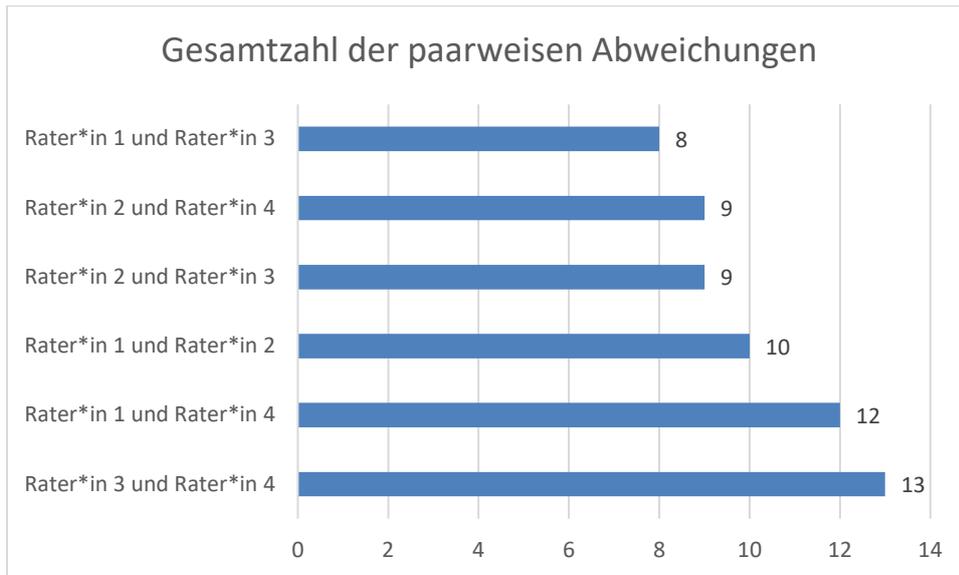


Abbildung 2: Gesamtzahl der paarweisen Abweichungen

Abbildung 3 zeigt nun diese Abweichungen nach Dimensionen und Paaren, wobei bei Dimension 1 keine Abweichungen bestehen und auch bei Dimension 3 stimmen die Bewertungen zumeist überein (s. auch Tab. 6).

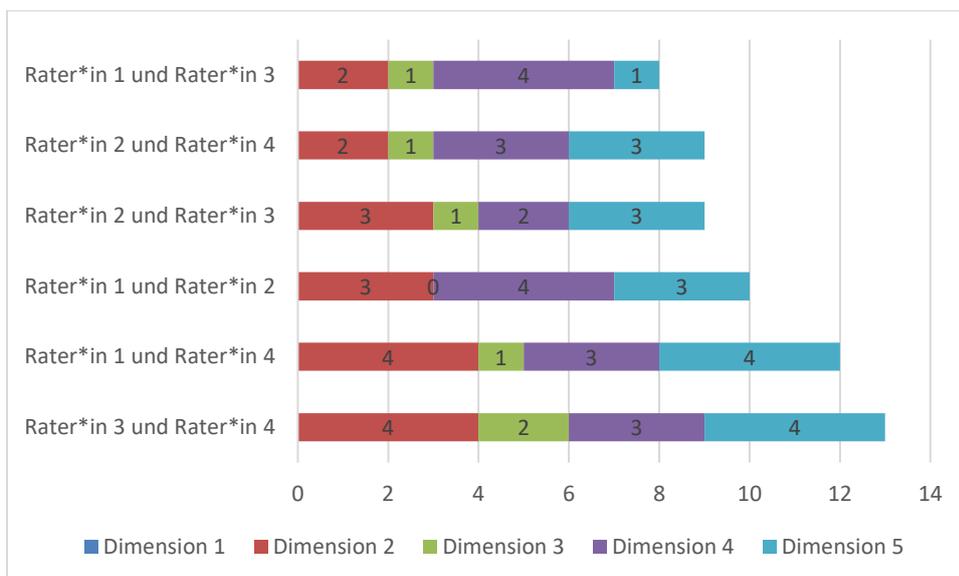


Abbildung 3: Gesamtzahl der paarweisen Abweichungen nach Dimensionen

	Rater*in 3 und Rater*in 4	Rater*in 1 und Rater*in 4	Rater*in 1 und Rater*in 2	Rater*in 2 und Rater*in 3	Rater*in 2 und Rater*in 4	Rater*in 1 und Rater*in 3
Dimension 1	0	0	0	0	0	0
Dimension 2	4	4	3	3	2	2
Dimension 3	2	1	0	1	1	1
Dimension 4	3	3	4	2	3	4
Dimension 5	4	4	3	3	3	1

*Tabelle 6: Anzahl der Abweichungen nach Dimensionen*

Aufgrund der Analysen kann geschlussfolgert werden, dass die zufriedenstellenden Cronbachs Alpha Werte (Gesamt 0. 869) durchaus eine Interrater-Reliabilität ergeben, die ein weiteres Rating ermöglichen.

## Literatur

- Katschnig, T., Geppert, C. (2017): Die Bestimmung der Interrater-Reliabilität in einem TheoPrax-Teilprojekt. Ein Beitrag zur Auswertung von 12 WIRTH-Interviews. Institut für Bildungswissenschaft der Universität Wien und Institut Fortbildung für PädagogInnen der KPH Wien/Krems. Mit Ergänzungen vom 8.2.2021. Elektronisch aufrufbar unter: <http://phaidra.univie.ac.at/o:1163074>
- Strobl, B. (2017): Zur praxisleitenden Bedeutung von Theorien und Konzepten aus der Sicht psychoanalytischer PsychotherapeutInnen. Eine empirische Untersuchung. Masterarbeit am Institut für Bildungswissenschaft der Universität Wien
- Strobl, B. (2019): Bericht über das Projekt „Der Einsatz des Wiener Interviews zur Identifizierung der mentalen Repräsentation praxisleitender Theorien (WIRTH). Zur Untersuchung des Theorie-Praxis-Verhältnisses aus der Sicht von Absolventinnen und Absolventen von Wiener Ausbildungen im Bereich der Psychotherapie und Psychagogik“, gefördert durch die Kulturabteilung MA 7 der Stadt Wien. Unpublizierter Forschungsbericht
- Strobl, B., Datler, W. (2020): Emotionen als Gegenstand des Nachdenkens und Sprechens über Praxissituationen. Anmerkungen zur Bedeutung von psychoanalytisch orientierten Aus- und Weiterbildungsprozessen für eine Dimension von psychosozialer Professionalität. In: Rauh, B., Welter, N., Franzmann, M., Magiera, K., Schramm, J. & Wilder, N. (Hrsg.): Emotion – Disziplinierung – Professionalisierung. Budrich: Opladen et al., 207-224
- Strobl, B., Datler, W. (2021): Psychotherapeutisch Tätige geben Auskunft: Über die Orientierung von Praxis an Theorie und den Einfluss psychoanalytischer Aus- und Weiterbildungsstrukturen. Erscheint voraussichtlich 2021 in: Psychotherapie Forum 25. Elektronisch abrufbar unter: (<https://www.springer.com/journal/729>)
- Wirtz, M., Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. München: Hogrefe