

Klaus D. Kubinger

**Psychologische
Diagnostik**

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

(Vortrag im Rahmen der Lehrveranstaltung: Anwendungsfehler in Statistik, Psychometrie und Diagnostik“)

„Warum der Einsatz der meisten psychologischen Tests zu Rechtsklagen bei der praktischen Fallbegutachtung führen kann“

Klaus D. Kubinger

Artikel eingereicht in *Report Psychologie*



Vermutlich würde jede(r) in der praktischen Fallbehandlung tätige Psychologe/in in Deutschland (und genauso in Österreich) mit viel Selbstbewusstsein behaupten, **psychologische Begutachtungen mit höchster, nämlich berufsständisch gebotener Sorgfalt** durchzuführen.

Und tatsächlich erfüllen wohl die meisten davon weitgehend die Anforderungen, wie sie zum Beispiel in den *Berufsethischen Richtlinien* des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. und der Deutschen Gesellschaft für Psychologie e.V. zugleich Berufsordnung des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. (in der von der Delegiertenkonferenz bzw. der Mitgliederversammlung 2016 beschlossenen Fassung) festgelegt sind.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Dort heißt es vor allem:

„Psychologinnen und Psychologen ... pflegen eine größtmögliche sachliche und wissenschaftliche Fundiertheit, **Sorgfalt** und Gewissenhaftigkeit bei der Erstellung und Verwendung von Gutachten und Untersuchungsberichten“.

- ▶ <https://www.bdp-verband.de/binaries/content/assets/beruf/ber-foederation-2016.pdf>
- ▶ in Österreich sind bestimmte Berufspflichten im *Psychologengesetz* (österreichisches BGBl.Nr. 182/2013) geregelt



Geht man allerdings etwas ins Detail, so lassen sich jedoch **Zweifel** an der Einhaltung der damit erforderlichen „Sorgfaltspflicht“ befürchten.

Heute geht es ausschließlich um die **Sorgfalt** bei der **Auswahl** der konkret eingesetzten *psychologisch-diagnostischen Verfahren* zur Begutachtung; und zwar auch nur soweit es **einschlägig** auf die **Verfahren** betrifft.

Verfahren, deren Quelle eine gewisse **Grauzone** darstellen und die nicht alle Entwicklungsschritte im Detail offengelegen bzw. nicht ausreichende Befunde zu den Gütekriterien berichten, sind von vornherein kritisch zu sehen.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Es interessiert also im Folgenden (nur), ob die Auswahl der konkreten Verfahren

zur Messung der fallrelevanten Eigenschaften der psychologisch zu untersuchenden Person ausreichend **sorgfältig** erfolgt.

Und das nimmt Bezug

zur **Messtheorie** im Allgemeinen

und zur **psychologischen Testtheorie** im Besonderen,

also der mathematisch-statistisch fundierten **Theorie des Messens psychischer Phänomene**.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Bekanntermaßen ergänzen diese Theorie im Sinne einer „**Lehre zur Testkonstruktion**“ bestimmte **Gütekriterien**, die entsprechende Messinstrumente erfüllen müssen.

Heutiges Thema:

die beiden Gütekriterien **Eichung** und **Skalierung** ...

... hinsichtlich der erforderlichen **Sorgfalt** bei einer psychologischen Begutachtung

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Ein Test erfüllt das Gütekriterium *Eichung*, wenn für sein Bezugssystem zur Relativierung des individuellen Testergebnisses (die sog. Eich Tabellen) folgende Bedingungen gegeben sind:

- Die Eich Tabellen sind gültig, d. h. nicht veraltet,
- die Population, für die die Eich Tabellen gelten, ist definiert,
- die für die Erstellung der Eich Tabellen herangezogene Stichprobe ist repräsentativ.

← Kubinger (2009, S. 68)



Ein Test erfüllt das Gütekriterium *Eichung*, wenn für sein Bezugssystem zur Relativierung des individuellen Testergebnisses (die sog. Eich Tabellen) folgende Bedingungen gegeben sind:

- Die Eich Tabellen sind gültig, d. h. nicht veraltet,
- die Population, für die die Eich Tabellen gelten, ist definiert,
- die für die Erstellung der Eich Tabellen herangezogene Stichprobe ist repräsentativ.

← Kubinger (2009, S. 68)

Besonders kritisch steht es oft um die Aktualität der Eich Tabellen.

Wegen des steten Wandels der heutigen Gesellschaft ist sie nach einigen Jahren immer zweifelhaft.

Zeigte zum Beispiel der so genannte Flynn-Effekt früher, dass der Durchschnittswert des IQ innerhalb von ca. 20 Jahren bis zu 25 Punkte ansteigt, so sprechen jüngere Ergebnisse für einen umgekehrten Trend (Dutton, van der Linden & Lynn, 2016).

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Ein Test erfüllt das Gütekriterium *Skalierung*, wenn die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden.

← Kubinger (2009, S. 82)

Es geht um die Angemessenheit der im Manual eines Tests festgesetzten Reglementierung, **wie die einzelnen Testleistungen** einer Testperson **zu einem numerischen Testwert zu verrechnen sind** – mit „Angemessenheit“ ist dabei gemeint, dass faktische Gegebenheiten entsprechend repräsentiert werden.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens

2., überarbeitete und erweiterte Auflage



HOGREFE

Didaktisch wohl nach wie vor **das überzeugendste empirische Beispiel** der Verletzung dieses Gütekriteriums:

Klaus D. Kubinger

www.klaus-kubinger.com



**universität
wien**

c/o

Fakultät für Psychologie
Arbeitsbereich für Psychologische Diagnostik

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE



**„Was mußt du tun, damit das
Wasser kocht?“**

www.aid-intelligenztest.at

klaus.kubinger@univie.ac.at



Didaktisch wohl nach wie vor **das überzeugendste empirische Beispiel** der Verletzung dieses Gütekriteriums:

„Was mußt du tun, damit das Wasser kocht?“ →

Kindern mit **niedrigem** Testwert („Weniglösern“), also leistungsschwachen Kindern, fällt diese Aufgabe relativ leichter als Kindern mit **hohem** Testwert („Viellösern“), also leistungsstarken Kindern.

Weil sogar die recht anspruchslose Antwort „**auf den Herd stellen**“ als richtig gilt, misst diese Aufgabe in hohem Maß (auch) das **Anspruchsniveau** der Tp: **Leistungsstarke Kinder/Jugendliche werden sich mit dieser Antwort nicht zufrieden geben**; eine physikalisch fundierte fällt jedoch auch ihnen schwer.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Didaktisch wohl nach wie vor **das überzeugendste empirische Beispiel** der Verletzung dieses Gütekriteriums:

„Was mußt du tun, damit das Wasser kocht?“ →

Kindern mit **niedrigem** Test-
leistung

Diese Aufgabe (wie alle anderen) bei „Lösung“ mit 1 Punkt zu verrechnen, spiegelt **nicht** die empirisch feststellbaren Verhaltensrelationen wider:

... (auch) das
... Leistungsstarke Kinder/Jugendliche
werden sich mit dieser Antwort nicht zufrieden geben; eine
physikalisch fundierte fällt jedoch auch ihnen schwer.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Didaktisch wohl nach wie vor **das überzeugendste empirische Beispiel** der Verletzung dieses Gütekriteriums:

„Was mußt du tun, damit das Wasser kocht?“ →

Kindern mit **niedrigem** Test-
leistung

Diese Aufgabe (wie alle anderen) bei „Lösung“ mit 1 Punkt zu verrechnen, spiegelt **nicht** die empirisch feststellbaren Verhaltensrelationen wider:
Verhaltensmäßig ist durch die anderen Aufgaben empirisch belegt, dass ein „Viellöser“ eine höhere Fähigkeit hat als ein „Weniglöser“; bei dieser Aufgabe wird jedoch (oft) „verkehrt“ verrechnet, nämlich **die geringere Fähigkeit mit einem Punkt honoriert, die höhere Fähigkeit ohne Punkt gewürdigt.** ... (auch) das ... Leistungsstarke Kinder/Jugendliche werden sich mit dieser Antwort nicht zufrieden geben; eine physikalisch fundierte fällt jedoch auch ihnen schwer.

Klaus D. Kubinger

www.klaus-kubinger.com



universität
wien

c/o

Fakultät für Psychologie
Arbeitsbereich für Psychologische Diagnostik

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Ein anderes Beispiel zur Verletzung des Gütekriteriums Skalierung

www.aid-intelligenztest.at

klaus.kubinger@univie.ac.at

Klaus D. Kubinger

**Psychologische
Diagnostik**

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

**„Welche Farbe erhält man, wenn
man die Farben Blau und Gelb
miteinander vermischt?“**

**„Wie viele Menschen gibt es auf
der Welt?“**



Ein anderes Beispiel zur Verletzung des Gütekriteriums Skalierung
„Welche Farbe erhält man, wenn man die Farben Blau und Gelb miteinander vermischt?“

„Wie viele Menschen gibt es auf der Welt?“ →

... die erste Aufgabe fällt **Mädchen** relativ leichter als **männlichen Kindern/Jugendlichen**; bei der zweiten Aufgabe ist es genau umgekehrt.

Das zeigt, dass der Test nicht eindimensional misst, sondern einzelne Aufgaben etwas ganz anderes als die intendierte Fähigkeit erfassen. **Überspitzt formuliert, messen sie (bloß) das Geschlecht.**



Ein ganz anderes, nicht-psychologisches Beispiel (aus Kubinger, Rasch & Yanagida, 2011, S. 60):

Ein Beispiel des Sports illustriert deutlich, wie **willkürlich** und **unsachgemäß**, also nicht- bzw. bloß vorwissenschaftlich, Skalen **im Alltag** festgelegt werden.

Im (alpinen) **Skiweltcup** werden bekanntlich die Bestplatzierten eines Rennens mit **Weltcuppunkten** belohnt, wobei die **Gesamtsumme erzielter Punkte** am Ende der Saison den **Sieger** bestimmt.



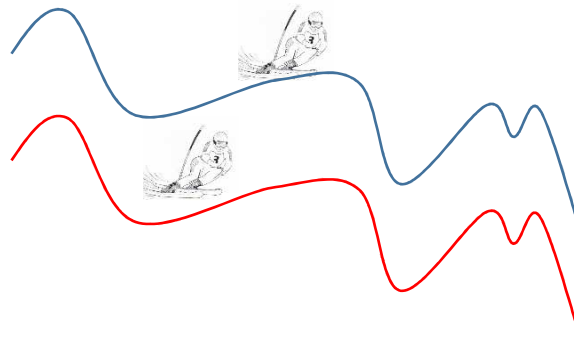
Nun wäre es ein Leichtes, die faktischen Leistungen etwa der beiden Läufer *R* und *S* gegenüberzustellen,

d. h. **die empirische Relation ihrer Schnelligkeiten** über die gesamte Saison durch Übereinanderlegen entsprechender Videoaufzeichnungen zu bestimmen (vorausgesetzt, beide haben dieselben Rennen bestritten).



Nun wäre es ein Leichtes, die faktischen Leistungen etwa der beiden Läufer *R* und *S* gegenüberzustellen,

d. h. **die empirische Relation ihrer Schnelligkeiten** über die gesamte Saison durch Übereinanderlegen entsprechender Videoaufzeichnungen zu bestimmen (vorausgesetzt, beide haben dieselben Rennen bestritten).



Das **Reglement** sieht jedoch eine andere Skalierung vor, **bildet** also **eine bestimmte numerische Relation ab**, die der empirischen widersprechen kann und so gesehen nicht fair ist:

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Der Sieger eines Rennens bekommt 100 Punkte,
der Zweite 80,
der Dritte 60,
der Vierte 50 usw.

Angenommen R ist bei 5 Rennen jedes Mal um 0,01s langsamer als der Gewinner S und immer Zweiter, im 6. Rennen jedoch Sieger und 2s schneller als S als Vierter.

Dann gewinnt S mit $(5 \times 100) + 50 = 550$ Punkten den Weltcup vor R mit $(5 \times 80) + 100 = 500$ Punkten,

obwohl das Übereinanderlegen der Videoaufzeichnungen erbringt, dass R insgesamt, über die Saison hinweg, 1,95s schneller war als S .

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Eine solche **unfaire Skalierung** ist allerdings **solange belanglos**, wie sich **Sportler** und **Zuschauer** aus Unterhaltungsgründen dem **autoritären Reglement** freiwillig unterwerfen –

beim **psychologischen Diagnostizieren** in Fallbehandlungen mit dem Anspruch an Fairness und Objektivität wäre eine **solche unfaire Skalierung jedoch ethisch nicht vertretbar!**



So **befremdlich** es selbst zu Zeiten von *Alfred Binet* oder gar *David Wechsler* war, sich nämlich **bei psychologischen Tests über die Erkenntnisse der Messtheorie** aus der Physik hinwegzusetzen, so ist es spätestens seit *Georg Rasch* in den 60er Jahren des letzten Jahrhunderts auch innerhalb der Psychologie völlig **unakzeptabel**, die **Verrechnungsmodi** von Tests **nicht empirisch abzusichern**.

Klaus D. Kubinger

**Psychologische
Diagnostik**

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Ausgangspunkt:

(leider wieder einmal) das sog. **Rasch-Modell!**

An dieser Stelle soll auf das *Rasch-Modell* gar nicht genauer eingegangen werden (s. eine kurze Einführung in *Dorsch – Lexikon der Psychologie*; Kubinger, 2017)



Wenn als Testwert **die Anzahl der gelösten Aufgaben** verrechnet wird

(d. h. **egal**, **welche Aufgaben eines Tests** **gelöst werden und welche nicht**, die Leistungsgüte einer Testperson wird allein durch die Anzahl gelöster Aufgaben bestimmt – zum Beispiel wird explizit nicht gewertet, inwieweit manche der gelösten Aufgaben schwieriger, andere leichter sind) ...

... und man also **voraussetzt**, **dass damit** die empirisch gegebenen (inter- und intraindividuellen)

Verhaltensrelationen adäquat **abgebildet werden**,

dann muss (*notwendigerweise!*) für alle Aufgaben eines psychologischen Tests das **Rasch-Modell** **gelten !!!**



Wenn als Testwert die Anzahl der gelösten Aufgaben
verrechnet wird, **dann**

muss das
Rasch-Modell
gelten

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



Wikipedia: Als empirische Wissenschaften oder Erfahrungswissenschaften gelten Disziplinen, in denen die Objekte und Sachverhalte der Welt, wie z. B. Planeten, Tiere, Verhaltensweisen von Menschen durch Experimente, Beobachtung oder Befragung untersucht werden.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

... „**notwendigerweise**“ entspringt einmal dem Umstand, dass die gemachte Aussage einem **mathematischen Beweis** zufolge stimmt (s. Fischer, 1974, 1995).

D. h., diese Aussage

[wenn als Testwert die Anzahl der gelösten Aufgaben verrechnet wird, dann muss für alle Aufgaben eines psychologischen Tests das Rasch-Modell gelten]

hat die Bedeutung **naturwissenschaftlicher Verbindlichkeit** (es gibt keinen Spielraum für Interpretation bzw. Relativierung)

~~... für meinen Test gilt zwar das Rasch-Modell nicht, aber die Praktiker(innen) wenden ihn so gerne an ...~~



... „**notwendigerweise**“ hat aber auch Folgendes zur Konsequenz:

Psychologische Tests, welche zwar den bewussten Verrechnungsmodus verwenden, aber dessen Aufgaben nachgewiesener Maßen **nicht** (alle) dem Rasch-Modell entsprechen, ...

... **bilden** die **empirisch feststellbaren Verhaltensrelationen** innerhalb einer Testperson bzw. zwischen verschiedenen Testpersonen einfach **inadäquat ab**; d. h.

sie führen zu realitätsverfälschenden Testwerten

Klaus D. Kubinger

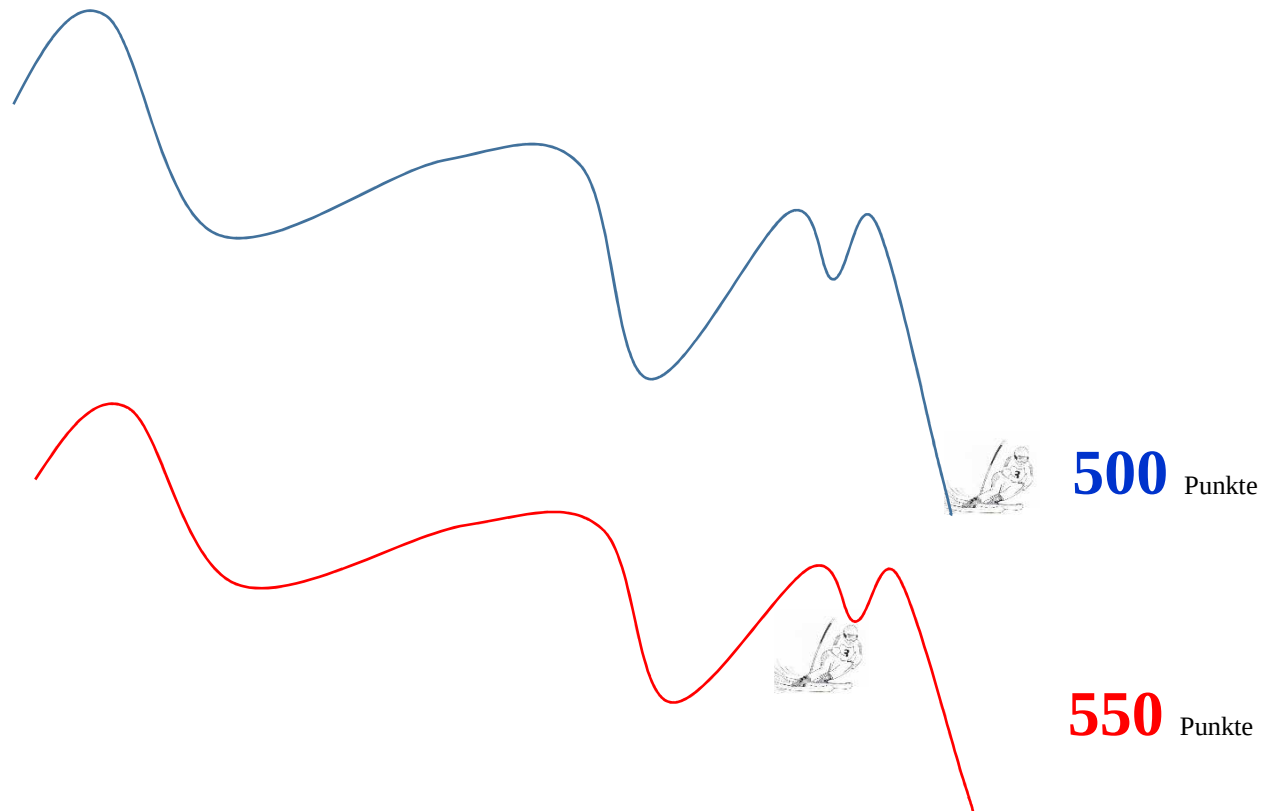
Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

... vergleiche:



Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Und wurde ein psychologischer Tests mit diesem Verrechnungsmodus gar nie hinsichtlich der Verträglichkeit (aller) seiner Aufgaben mit dem Rasch-Modell geprüft, ...

dann ist jedenfalls zu **befürchten**, dass er die empirisch feststellbaren Verhaltensrelationen **innerhalb einer Testperson** bzw. **zwischen verschiedenen Testpersonen** **nicht adäquat** durch die erhaltenen Testwerten widerspiegelt.

Tatsächlich erwiesen sich bisher alle Tests, die (noch) ohne Rücksicht auf den genannten Beweis entwickelt wurden, hinsichtlich ihrer Verrechnung als unangemessen: Das *Rasch-Modell* gilt nicht!

Das macht es extrem **unplausibel**,
dass Tests, die nicht gleich entsprechend „kalibriert“ wurden,
trotzdem, quasi glücklicherweise, verrechnungsfair sind.

Klaus D. Kubinger

Psychologische Diagnostik

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Heute nicht:

Welches andere Modell (der sog. Item-Response-Theorie; IRT) mit welchem anderen Verrechnungsmodus psychologischer Tests korrespondiert.

Weil:

Ohnehin im Inventar psychologischer Tests hauptsächlich dieser Verrechnungsmodus vorkommt.

Um z.B. nur Intelligenz-Testbatterien anzuführen:

IST 2000 R, WIT-2, WISC-V, WAIS-IV, ids-2, KABC-II, **AID 3**



Zwar erfüllt keines analog dazu die „**Notwendigkeitsbedingung**“ – zur Angemessenheit eines bestimmten Verrechnungsmodus‘ müsse (!) ein ganz bestimmtes Modell gelten –;

immerhin ist aber für einige Modelle deren **Geltung hinreichend**.

D. h., wenn für einen psychologischen Test das fragliche Modell gilt bzw. dieses die beobachteten Daten ausreichend gut erklärt, dann ist der entsprechende Verrechnungsmodus auch angemessen.

Gilt es jedoch nicht bzw. erklärt es die beobachteten Daten nicht ausreichend gut, dann ist eben nicht gewährleistet, dass die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen tatsächlich adäquat beschreiben.

Klaus D. Kubinger

**Psychologische
Diagnostik**

Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



HOGREFE

Rechtsgrundlage

→ Grundsätzlich

„können **berufsethische Grundsätze**, die in der Berufsordnung enthalten sind, Ausdruck einer einheitlichen und gefestigten Standesüberzeugung sein, die auch von der Allgemeinheit geteilt wird, und **deren Verletzung allgemein als Verstoß gegen das Anstandsgefühl angesehen wird**“
(Fehnmann, 1995, S. 133).



Rechtsgrundlage

Erst recht gilt das betreffs der *Qualitätsstandards für psychologische Gutachten* (Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen, 2017) und der *Richtlinie für die Erstellung von klinisch-psychologischen und gesundheitspsychologischen Befunden und Gutachten* vom „Psychologenbeirat“ des österreichischen BM für Gesundheit.



Rechtsgrundlage

Leider nehmen beide, „Qualitätsstandards“ und „Richtlinie“, **nicht** Bezug auf die Gütekriterien **Eichung** und **Skalierung**.

Allerdings tut dies direkt oder zumindest indirekt die

DIN 33430 (DIN Deutsches Institut für Normung e.V., 2002; bzw. 2016) –

Qualitätsstandards von Normungsinstituten stellen **oft die rechtsnächste Norm** dar, was heißt, **sie werden im Rechtswesen regelmäßig eingesetzt**.



Rechtsgrundlage

*Als Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen bzw. Anforderungen an berufsbezogene Eignungsdiagnostik verlangt die **DIN***

33430 nun explizit, **dass**

die „Angemessenheit der Normwerte in den letzten acht Jahren überprüft worden“ ist (2016; S. 29).

Das schließt definitiv die Anwendung von Verfahren aus, deren Bezugssystem zur Relativierung eines individuellen Testergebnisses **überaltert und nicht einigermassen aktuell bestätigt sind.**



„Die in der DIN 33430 geforderte Maximalanzahl von acht Jahren bis zur Gültigkeitsüberprüfung der Eichtabellen eines Tests erscheint in Anbetracht des dafür nötigen Aufwands etwas überzogen. Da die Festsetzung dieser Größe völlig willkürlich erfolgte, nämlich ohne dass es irgendwelche sachbezogenen Argumente gibt, warum nicht statt acht Jahren sechs oder sieben oder neun oder zehn Jahre gewählt worden sind, sollte die Anzahl von acht Jahren besser nur als grober Richtwert aufgefasst werden. Selbstverständlich zielt aber die gegebene Forderung darauf ab, nicht Tests einzusetzen, deren Eichstichprobenerhebung zwei Jahrzehnte oder noch mehr zurück liegt.“ (Kubinger, 2019)



Rechtsgrundlage

Indirekt bezieht sich die DIN 33430 auch auf die **Skalierung**.

Aus ihren Anforderungen wurden **Richtlinien für die Beurteilung von psychologisch-diagnostischen Verfahren**

zur Erfassung menschlichen Erlebens und Verhaltens
abgeleitet (Diagnostik- und Testkuratorium, 2018a, b);

und dieses „Testbeurteilungssystem“ fordert nachdrücklich
die Berücksichtigung des Gütekriteriums **Skalierung**:

„Insbesondere die IRT, d. i. vor allem das Rasch-Modell, bringt es mit sich, dass bei Tests auch kritisch hinterfragt wird, inwieweit die Zahlenrelationen der Testwerte mit den Relationen der beobachtbaren Verhaltensweisen – sowohl innerhalb ein und derselben Testperson als auch zwischen verschiedenen Testpersonen – übereinstimmen (‚Skalierung‘).

Da eine entsprechende empirische Absicherung durch die Testautorinnen und -autoren eben nur durch den Einsatz der Modelle der IRT möglich ist, sollten die Rezensentinnen und Rezensenten nicht nur eine gegebenenfalls versuchte Absicherung dieser Art beurteilen, sondern auch im Fall, dass die Testkonstruktion nicht nach der

erfolgte, wenigstens anführen, dass es danach bei der Beurteilung von Tests genügt, das bewusste Thema lediglich aufzugreifen und zu diskutieren, eröffnet vordergründig zwar eine gewisse Unverbindlichkeit; **letztlich wird aber doch klargestellt, dass die Verrechnungsvorschrift eines Tests empirisch abgesichert werden muss.**

enden.“ (S. 113/S. 116).



Rechtsgrundlage

Was die Auswahl der *psychologisch-diagnostischen Verfahren* für eine Begutachtung betrifft, steht damit die Prüfung der Einhaltung der **Sorgfaltspflicht** des/r **fallführenden Psychologen/in** auch in Bezug auf die Gütekriterien Eichung und Skalierung nicht im rechtsfreien Raum.

§



Rechtsgrundlage

Noch ist zwar fraglich, ob es diesbezüglich tatsächlich zu Rechtsklagen kommt; sich also Klienten unsachgemäß begutachtet fühlen, und auf Richtigstellung oder gar Schadenersatz klagen.

Nach BDP-Justiziar Rechtsanwalt **Jan Frederichs**:



Rechtsgrundlage

... würden Gerichte dann aus juristischer Sicht den „Sorgfaltsmaßstab“ bestimmen:

D. h., sie legen im konkreten Streitfall fest, welche Aussagen des hinzugezogenem Sachverständigen den **Fachstandard** darstellen und welche Aussagen seine fachlich subjektive Meinung über die *best Practice*, die eher in der Fachwelt (noch) nicht den Rang eines anerkannten Standards erreicht hat.



So gesehen scheint es in der Praxis durchaus **riskant**, solche *psychologisch-diagnostische Verfahren* einzusetzen, die den Gütekriterien Eichung und/oder Skalierung nach wissenschaftlichen Maßstäben nicht (mehr) entsprechen:

Als Anwender/in kann man nur hoffen, dass gegebenenfalls das Gericht entscheidet, diese Maßstäbe gehen über den allgemeinen Fachmeinungsstandard hinaus.



Werden zur Erstellung eines psychologischen Gutachtens nur solche *psychologisch-diagnostischen Verfahren* verwendet, deren Eichung als hinreichend aktuell gelten (die Gültigkeitsüberprüfung ist nicht wesentlich älter als acht Jahre) und deren Skalierung abgesichert wurde (die Verrechnungsvorschriften bilden die empirischen Verhaltensrelationen adäquat ab), so ist **der/die Gutachter(in) auf der sicheren Seite**: Es besteht dann **kein Anlass zur Rechtsklage** auf entsprechende **Verletzung der Sorgfaltspflicht**.

Klaus D. Kubinger

www.klaus-kubinger.com



**universität
wien**

c/o

Fakultät für Psychologie
Arbeitsbereich für Psychologische Diagnostik

**Vielen Dank für Ihre
Aufmerksamkeit!**

Klaus D. Kubinger

**Psychologische
Diagnostik**
Theorie und Praxis
psychologischen Diagnostizierens
2., überarbeitete und erweiterte Auflage



3. Aufl., 2019,
in Druck

