



DIE AUSWIRKUNG DER „VERORDNUNG (EU) 2017/745 DES EUROPÄISCHEN PARLAMENTS UND DES RATES AUF MEDIZINPRODUKTE“ AUF DIE BEWERTUNG MESSTHEORETISCH FUNDIERTER TESTS

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen
verabschiedet am 3. Februar 2022

Vorwort

Dieses Dokument wurde im Auftrag der Föderation Deutscher Psychologinnenvereinigungen vom Diagnostik- und Testkuratorium (DTK) erstellt. Mitglieder des Diagnostik- und Testkuratoriums zum Zeitpunkt der Texterstellung waren: Prof. Dr. Markus Bühner, Dr. Tom Frenzel, Prof. Dr. Carmen Hagemeister, Prof. Dr. Nina Heinrichs, Prof. Dr. Martin Kersting (Vorsitzender), Dipl.-Psych. Fredi Lang, Dr. Gerd Reimann und Prof. Dr. Matthias Ziegler.

Die folgende Stellungnahme zur Auswirkung der „Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates über Medizinprodukte“ auf die Bewertung messtheoretisch fundierter Tests wurde vom Diagnostik- und Testkuratorium (DTK) verfasst. Das DTK ist ein von der „Föderation Deutscher Psychologinnenvereinigungen“ getragenes Gremium, das u.a. zuständig ist für alle Aspekte der Qualitätssicherung und Qualitätsoptimierung des diagnostischen Prozesses in Forschung und Anwendung, soweit es sich um Diagnostik menschlichen Erlebens und Verhaltens handelt. Dazu gehört insbesondere die Qualitätssicherung diagnostischer Verfahren und die Information und Aufklärung der Öffentlichkeit über Möglichkeiten und Grenzen psychologischer Diagnostik.

In dieser Stellungnahme wird zunächst dargelegt, wie Medizinprodukte gemäß der Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates über Medizinprodukte (im Folgenden „Verordnung (EU) 2017/745“) definiert sind und nach welchen Kriterien sie weiter unterteilt werden. Darüber hinaus wird kurz dargestellt, wie im Rahmen der Verordnung (EU) 2017/745 eine Qualitätssicherung zu erfolgen hat. Vergleichend erfolgt danach eine entsprechende Darstellung für Tests. Aufbauend auf diesen Angaben erfolgt eine Empfehlung bezüglich der Bewertung von messtheoretisch fundierten Tests in Bezug auf Verordnung (EU) 2017/745. Abschließend werden mögliche Konsequenzen besprochen.

Die Stellungnahme richtet sich insbesondere an Hersteller*innen und Nutzer*innen von messtheoretisch fundierten Tests, sowie Instanzen, welche im Rahmen der Umsetzung von Verordnung (EU) 2017/745 mit der Einstufung von Verfahren als Medizinprodukte betraut sind. Schließlich richtet sich die Stellungnahme auch an politische Akteur*innen.

●● Föderation Deutscher Psychologenvereinigungen

1. Anlass

Am 25. Mai 2017 trat die Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates über Medizinprodukte in Kraft. Aufgrund der Corona-Pandemie wurde der Geltungsbeginn auf den 26. Mai 2021 verschoben. Die zahlreichen Änderungen in Bezug auf vorherige Verordnungen beinhalten auch neue Klassifizierungsregeln für Software, welche im Rahmen klinischer Diagnostik eingesetzt wird. Damit besteht zumindest theoretisch erst einmal eine Implikation für softwarebasierte, messtheoretisch fundierte Tests, welche im Rahmen klinischer Diagnostik eingesetzt werden. Die Verordnung in verschiedenen Sprachen findet sich hier: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02017R0745-20200424>

2. Was ist ein Medizinprodukt?

Im Folgenden werden relevante Auszüge der Verordnung (EU) 2017/745 aufgelistet und wichtige Passagen durch Fettdruck hervorgehoben.

Gemäß Artikel 2, Ziffer 1 der Verordnung (EU) 2017/745 gilt:

„Medizinprodukt' bezeichnet ein Instrument, einen Apparat, **ein Gerät, eine Software**, ein Implantat, ein Reagenz, ein Material oder einen anderen Gegenstand, **das dem Hersteller zufolge für Menschen bestimmt ist** und allein oder in Kombination einen oder mehrere der folgenden spezifischen medizinischen Zwecke erfüllen soll:

- **Diagnose**, Verhütung, Überwachung, Vorhersage, **Prognose**, Behandlung oder Linderung **von Krankheiten**,
- Diagnose, Überwachung, Behandlung, Linderung von oder Kompensierung von Verletzungen oder Behinderungen,
- Untersuchung, Ersatz oder Veränderung der Anatomie oder eines physiologischen oder pathologischen Vorgangs oder Zustands,
- Gewinnung von Informationen durch die In-vitro-Untersuchung von aus dem menschlichen Körper – auch aus Organ-, Blut- und Gewebespenden – stammenden Proben

und dessen bestimmungsgemäße Hauptwirkung im oder am menschlichen Körper weder durch pharmakologische oder immunologische Mittel noch metabolisch erreicht wird, dessen Wirkungsweise aber durch solche Mittel unterstützt werden kann.“

Artikel 51 (1) regelt, dass Medizinprodukte in die Klassen I, IIa, IIb und III gemäß Anhang VIII eingestuft werden. Dort ist unter 4.1 Regel 1 definiert:

„Alle nicht invasiven Produkte gehören zur Klasse I, es sei denn, es findet eine der folgenden Regeln Anwendung.“

In Regel 6.3 findet sich:

„Software, die dazu bestimmt ist, Informationen zu liefern, die zu Entscheidungen für diagnostische oder therapeutische Zwecke herangezogen werden, gehört zur Klasse IIa, es sei denn, diese Entscheidungen haben Auswirkungen, die Folgendes verursachen können:

- den Tod oder eine irreversible Verschlechterung des Gesundheitszustands einer Person; in diesem Fall wird sie der Klasse III zugeordnet, oder
- eine schwerwiegende Verschlechterung des Gesundheitszustands einer Person oder einen chirurgischen Eingriff; in diesem Fall wird sie der Klasse IIb zugeordnet.

Software, die für die Überwachung von physiologischen Prozessen bestimmt ist, gehört zur Klasse IIa, es sei denn, sie ist für die Kontrolle von vitalen physiologischen

●● Föderation Deutscher Psychologenvereinigungen

Parametern bestimmt, wobei die Art der Änderung dieser Parameter zu einer unmittelbaren Gefahr für den Patienten führen könnte; in diesem Fall wird sie der Klasse IIb zugeordnet.

Sämtliche andere Software wird der Klasse I zugeordnet.“

Bezüglich der Anforderungen an die Qualität, welche für Medizinprodukte nachzuweisen sind, findet sich in Anhang I unter 15.1:

„Diagnostische Produkte und Produkte mit Messfunktion werden so ausgelegt und hergestellt, dass **auf der Grundlage geeigneter wissenschaftlicher und technischer Verfahren ausreichende Genauigkeit, Präzision und Stabilität für die Zweckbestimmung des Produkts** gewährleistet sind. Der Hersteller gibt die Genauigkeitsgrenzen an.“

Von Bedeutung dürfte auch 15.2 sein:

„**Die mithilfe von Produkten mit Messfunktion erstellten Messungen werden in gesetzlichen Einheiten entsprechend den Bestimmungen der Richtlinie 80/181/EWG des Rates ausgedrückt.**“

Diese Richtlinie (<https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32009L0003&from=EN>) beinhaltet folgende Tabelle:

1.2.3. Besondere Namen und Einheitenzeichen für abgeleitete SI-Einheiten

Größe	Einheit		Ausgedrückt	
	Name	Einheitenzeichen	in anderen SI-Einheiten	in den SI-Basiseinheiten
Ebener Winkel	Radian	rad		$\text{m} \cdot \text{m}^{-1}$
Räumlicher Winkel	Steradian	sr		$\text{m}^2 \cdot \text{m}^{-2}$
Frequenz	Hertz	Hz		s^{-1}
Kraft	Newton	N		$\text{m} \cdot \text{kg} \cdot \text{s}^{-2}$
Druck, mechanische Spannung	Pascal	Pa	$\text{N} \cdot \text{m}^{-2}$	$\text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-2}$
Energie, Arbeit, Wärmemenge	Joule	J	$\text{N} \cdot \text{m}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2}$
Leistung ^(\dot{Q}) , Energiefluss	Watt	W	$\text{J} \cdot \text{s}^{-1}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3}$
Elektrizitätsmenge, elektrische Ladung	Coulomb	C		$\text{s} \cdot \text{A}$
Elektrische Spannung, elektrische Potentialdifferenz, elektromotorische Kraft	Volt	V	$\text{W} \cdot \text{A}^{-1}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-1}$
Elektrischer Widerstand	Ohm	Ω	$\text{V} \cdot \text{A}^{-1}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-2}$
Leitwert	Siemens	S	$\text{A} \cdot \text{V}^{-1}$	$\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^3 \cdot \text{A}^2$
Kapazität	Farad	F	$\text{C} \cdot \text{V}^{-1}$	$\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^4 \cdot \text{A}^2$
Magnetischer Fluss	Weber	Wb	$\text{V} \cdot \text{s}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$
Magnetische Flussdichte	Tesla	T	$\text{Wb} \cdot \text{m}^{-2}$	$\text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$
Induktivität	Henry	H	$\text{Wb} \cdot \text{A}^{-1}$	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-2}$
Lichtstrom	Lumen	lm	$\text{cd} \cdot \text{sr}$	cd
Beleuchtungsstärke	Lux	lx	$\text{lm} \cdot \text{m}^{-2}$	$\text{m}^{-2} \cdot \text{cd}$
Aktivität (ionisierende Strahlung)	Becquerel	Bq		s^{-1}
Energiedosis, spezifische Energie, Kerma, Energiedosisindex	Gray	Gy	$\text{J} \cdot \text{kg}^{-1}$	$\text{m}^2 \cdot \text{s}^{-2}$
Äquivalentdosis	Sievert	Sv	$\text{J} \cdot \text{kg}^{-1}$	$\text{m}^2 \cdot \text{s}^{-2}$
Katalytische Aktivität	Katal	kat		$\text{mol} \cdot \text{s}^{-1}$

●● Föderation Deutscher Psychologenvereinigungen

In Artikel 52 (6) steht:

„Hersteller von Produkten der Klasse IIa, ausgenommen Sonderanfertigungen oder Prüfprodukte, unterliegen einer Konformitätsbewertung gemäß Anhang IX Kapitel I und III sowie einer Bewertung der technischen Dokumentation – gemäß Abschnitt 4 jenes Anhangs – zumindest eines repräsentativen Produkts jeder Produktkategorie.“

In diesen weiteren Ausführungen sind umfangreiche Qualitätssicherungsmaßnahmen vorgesehen, welche auch regelmäßige Berichte auf Basis eines Überwachungsplans beinhalten.

Zusammenfassend können folgende Punkte als besonders relevant erachtet werden:

- Die Hersteller*innen trägt mit seinen Verwendungshinweisen entscheidend zur Einstufung als Medizinprodukt bei.
- Software, welche automatisiert Messergebnisse erzeugt, die automatisch zur Diagnose beitragen, fallen in die Kategorie MP IIa.
- Die Qualität ist durch einen umfangreichen Prozess zu sichern. Der technische Fokus liegt dabei auf Genauigkeit, Präzision und Stabilität. In diesem Zusammenhang dürfte auch der Verweis auf die zu berücksichtigenden Maßeinheiten von Relevanz sein.

Hieraus leitet sich die Frage ab, ob messtheoretisch fundierte Tests, welche unter Nutzung von Software zur Diagnose beitragen, als Medizinprodukte und dann folgerichtig als Medizinprodukte der Kategorie IIa zu bewerten sind. Um dies zu erörtern, wird im Folgenden zunächst noch einmal dargestellt, was in der Psychologie unter einem messtheoretisch fundierten Test verstanden wird.

3. Was ist ein messtheoretisch fundierter Test?

Im Testbewertungssystem des Board of Assessments der European Federation of Psychological Associations (EFPA Board of Assessment, 2013) sind Tests so definiert:

“Following the Standards for Educational and Psychological Testing the label test is used for any ... evaluative device or procedure in which a sample of examinee’s behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardized process” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 3). Therefore, this review model applies to all instruments that are covered under this definition, whether called a scale, questionnaire, projective technique, or whatever.”

Ausschlaggebend ist hier der Aspekt der Messung gemäß einem standardisierten Prozess. Diese Messung erfolgt durch das Zuweisen von Zahlen zu Beobachtungen gemäß bestimmter Regeln (Stevens, 1946). Wichtig zu betonen ist, dass diese Zuweisung von Zahlen noch keine interpretierbare Größe ergibt, da psychometrische Messungen keine natürliche Einheit haben (Michell, 2001). Vielmehr erfolgt die Interpretation der Messung durch die Bezugnahme auf einen Vergleichswert (z. B. durch das Überführen in einen Normwert wie IQ beim normorientierten Testen oder den Vergleich mit einem Cut-off-Wert, der aus einer Vergleichsstichprobe abgeleitet wurde beim kriterienorientierten Testen). Diese Interpretationen haben also keine Maßeinheit, die zu den in der Richtlinie 80/181/EWG aufgeführten Maßeinheiten gehört.

Messtheoretisch fundierte Tests liefern zwar Informationen, welche im Rahmen der klinisch-psychologischen Diagnostik Berücksichtigung finden. Dies erfolgt jedoch nur nach Interpretation durch Diagnostiker*innen und unter Berücksichtigung weiterer Information (Diagnostik-

●● Föderation Deutscher Psychologenvereinigungen

und Testkuratorium der Föderation Deutscher Psychologenvereinigungen, 2017; Witteman, Van der Heijden, & Claes, 2018).

Zur Beurteilung der Qualität von messtheoretisch fundierten Tests existieren eine Reihe nationaler (Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen, 2018) sowie internationaler Richtlinien (AERA, APA, & NMCE, 2014; EFPA Board of Assessment, 2013; Evers, Sijtsma, Lucassen, & Meijer, 2010). All diese Richtlinien haben gemein, dass sie die Qualität auf Basis verschiedener Kriterien definieren. Zu diesen gehören neben der Reliabilität (Messgenauigkeit, Präzision, Stabilität) auch die Objektivität, Validität (Gültigkeit), Normgruppe, Fairness, etc. Damit ist die Qualitätsprüfung wesentlich breiter aufgestellt als dies die Verordnung (EU) 2017/745 für Medizinprodukte vorsieht. Auch die Detailtiefe ist in den Richtlinien wesentlich größer. So finden sich in der Regel genaue Checklisten mit den in Bezug auf die Qualitätskriterien zu prüfenden Angaben.

Mit Blick auf die Verordnung (EU) 2017/745 ist die Umsetzung von messtheoretisch fundierten Tests in Software relevant. Dabei ist zu beachten, dass es in der Regel keinen funktionalen Unterschied zwischen einem mittels Papier und Stift durchgeführten Test und einem computerbasierten Test in Bezug auf die relevante Passage der Verordnung (EU) 2017/745 gibt. In beiden Fällen erfolgt die Messung durch das Zuweisen von Zahlen. Softwarebasiert geschieht dies computerbasiert anstatt mit Auswertungsschablonen. Die Überführung in Normwerte erfolgt ebenfalls computerbasiert, aber mittels derselben Vorschriften wie durch menschliche Testleiter*innen angewandt. So kann durch eine Softwarebasierung die Fehlerrate minimiert werden. Der entstehende Bericht, sprich die Information, ist jedoch gleich. Deshalb kommt der Softwarenutzung keine besondere diagnostische Bedeutung zu. Es wäre widersinnig, ein und dasselbe Instrument einmal als Medizinprodukt zu klassifizieren (computergestütztes Vorgehen) und einmal nicht („händisches“ Vorgehen).

4. Ist ein messtheoretisch fundierter Test ein Medizinprodukt?

Auf Basis der zuvor bereit gestellten Informationen wird hier geschlussfolgert, dass messtheoretisch fundierte Tests keine Medizinprodukte sind. Dies ist auch zutreffend für den Fall, dass die Verfahren softwarebasiert eingesetzt werden.

Entscheidend für diese Bewertung ist, dass messtheoretisch fundierte Tests zwar softwarebasiert eingesetzt werden können, aber eben keine eigene Software darstellen. Andernfalls wäre jeder digital administrierte Patientenfragebogen, der ja auch Angaben abfragen kann, welche bei der Diagnostik berücksichtigt werden (z. B. Alter, Geschlecht, Drogenkonsum), ein Medizinprodukt der Kategorie IIa mit den entsprechend notwendigen Qualitätssicherungsmaßnahmen. Auch ein Textverarbeitungsprogramm, mit dem ein Befund erstellt oder eine Krankenakte diagnostisch gefüllt wird, leistet einen Beitrag zur Erstellung einer Diagnose, aber nicht im Sinne einer technischen Messung, deren prozessuale Qualität gesichert werden müsste. Ein rechtlich relevanter Beitrag einer Software sollte in der Interpretation der Regelung daher eigenständiger sowie substanzieller und eher technischer Natur sein.

Weiterhin wird festgestellt, dass der Fokus der Qualitätssicherung bei Medizinprodukten auf der Messgenauigkeit liegt. Dies ist berechtigt, da die Messungen gemäß Richtlinie 80/181/EWG natürliche Einheiten haben, welche unmittelbar interpretierbar sind. Bei messtheoretisch fundierten Tests ist dies nicht möglich, die verwendeten Werte tauchen demnach auch nicht in der Richtlinie 80/181/EWG auf. Eine Interpretation erfolgt durch den Vergleich mit einer empirisch gewonnenen Vergleichsstichprobe. Dies gilt auch bei neuropsychologischen Verfahren, die beispielsweise Reaktionszeiten nutzen. Aus diesem Grund ist der Nachweis der Validität dieser Testwertinterpretation für Tests von herausgehobener Bedeutung. Entsprechend

●● Föderation Deutscher Psychologinnenvereinigungen

legen die oben genannten Richtlinien zur Bewertung der Qualität von messtheoretisch fundierten Tests viel Wert auf die Prüfung der Validität: Misst das Verfahren das, was es zu messen behauptet? Dieser Unterschied zwischen einer Messung im medizinischen versus psychologischen Sinn ist von erheblicher Bedeutung und stützt die Aussage, messtheoretisch fundierte Tests nicht als Medizinprodukte einzustufen.

5. Empfehlungen und Konsequenzen

Es wird klar empfohlen, messtheoretisch fundierte Tests nicht als Medizinprodukte einzustufen, auch, wenn diese softwarebasiert administriert werden.

Aus Sicht des DTK ist hervorzuheben, dass die Qualitätssicherung für messtheoretisch fundierte Tests aufgrund der Unterschiede in der Messung deutlich breitere Schwerpunkte setzt. Eine Klassifizierung als Medizinprodukt könnte dazu führen, dass die Qualität messtheoretisch fundierter Tests durch eine Fokussierung auf Messgenauigkeit und eine Vernachlässigung der Validitätsbelege perspektivisch sinken würde. Dies hätte potenziell dramatische Auswirkungen für die Güte der Diagnosen. Die Begutachtung der Qualität messtheoretisch fundierter Tests sollte nach Ansicht des DTK durch Psycholog(inn)en auf der Basis einschlägiger Standards erfolgen, die von psychologischen Fachorganisationen herausgegeben wurden.

6. Literatur

- AERA, APA, & NMCE. (2014). *Standards for Educational & Psychological Testing*. Washington, DC, USA.
- Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen. (2017). *Qualitätsstandards für Psychologische Gutachten*.
- Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen. (2018). Testbeurteilungssystem des Diagnostik und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. *Report Psychologie*, 43, 106–113.
- EFPA Board of Assessment. (2013). EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests. Retrieved from <http://www.efpa.eu/professional-development/assessment>
- Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch Review Process for Evaluating the Quality of Psychological Tests: History, Procedure, and Results. *International Journal of Testing*, 10, 295–317.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, 36, 211–218.
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 677–680.
- Witteman, C. L. M., Van der Heijden, P., & Claes, C. (2018). *Clinical assessment: Psychodiagnostic decision making*. Utrecht, NL: De Tijdstroom.

Hinweis: Bitte zitieren Sie diesen Text wie folgt:

Diagnostik- und Testkuratorium. (2022). *Stellungnahme des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen (DTK) zur Auswirkung der „Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates über Medizinprodukte“ auf die Bewertung messtheoretisch fundierter Tests*. Berlin: Föderation Deutscher Psychologinnenvereinigungen.

●● Föderation Deutscher Psychologinnenvereinigungen



Diese Arbeit ist lizenziert unter eine Creative-Commons-Lizenz CC BY-ND: Sie dürfen das Werk – solange Sie die Urheberin/den Urheber nennen – kopieren und für Ihre Zwecke nutzen. Dabei ist egal, ob die Nutzung kommerziell ist oder nicht. Allerdings darf das Werk nicht verändert werden.