IMPACT OF THE "REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL ON MEDICAL DEVICES" ON THE EVALUATION OF TESTS BASED ON MEASUREMENT THEORY

Statement from the Board of Assessment (January 25, 2023)

EFPA, The European Federation of Psychologists‘ Associations (established 1981) has the mission to develop and apply psychology for a positive impact on European society and beyond. EFPA publications are regularly consulted to inform EU policy and process. Now consisting of 37 European country associations, EFPA represents almost half of the world’s Psychologists whose members are required to observe professional standards. Specifically, EFPA Board of Assessment, whose members have approved this statement, convenes regularly to encourage and advance best practices in testing and assessment.

The following statement on the impact of the "Regulation (EU) 2017/745 of the European Parliament and of the Council on medical devices" on the evaluation of tests based on measurement theory originated from the Board of Assessment and Testing (DTK). The DTK is a body of the "Federation of German Psychological Psychologists’ Associations", which is responsible, among other things, for all aspects of quality assurance and quality optimization of the diagnostic process in research and application, as far as diagnostics of human experience and behavior are concerned. This includes in particular the quality assurance of psychological diagnostic procedures and the information and education of the public about the possibilities and limitations of psychological diagnostics.

This statement first outlines how medical devices are defined according to Regulation (EU) 2017/745 of the European Parliament and of the Council concerning medical devices (hereinafter "Regulation (EU) 2017/745") and according to which criteria they are further subdivided. In addition, it is briefly described how quality assurance must be carried out within the framework of Regulation (EU) 2017/745. This is followed by a comparative presentation for psychometric tests. Based on this information, a recommendation is made regarding the evaluation of psychometric tests based on measurement theory in relation to Regulation (EU) 2017/745. Finally, possible consequences are discussed.
The statement is aimed in particular at manufacturers and users of psychometric tests, as well as bodies that are entrusted with the classification of procedures as medical devices in the context of the implementation of Regulation (EU) 2017/745. Finally, the statement is also addressed to political actors.

Reason for this statement

On May 25, 2017, Regulation (EU) 2017/745 of the European Parliament and of the Council concerning medical devices entered into force. Due to the Corona pandemic, the effective date was postponed to May 26, 2021. The numerous changes to previous regulations also include new classification rules for software used in clinical diagnostics. Thus, at least theoretically, there is an implication for software-based, psychometric tests used in clinical diagnostics. The regulation in different languages can be found here.

What is a medical device?

In the following, relevant excerpts of Regulation (EU) 2017/745 are listed and important passages are highlighted in bold.

According to Article 2, Number 1 of Regulation (EU) 2017/745 applies:

"medical device' means any **instrument**, apparatus, appliance, **software**, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, **for human beings** for one or more of the following specific medical purposes:

- **diagnosis**, prevention, monitoring, prediction, **prognosis**, treatment or alleviation **of disease**,
- diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,
- investigation, replacement or modification of the anatomy or of a physiological or pathological process or state,
- providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations, and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means.

The following products shall also be deemed to be medical devices:

- devices for the control or support of conception;

- products specifically intended for the cleaning, disinfection or sterilisation of devices as referred to in Article 1(4) and of those referred to in the first paragraph of this point."

Article 51 (1) regulates that medical devices are classified in classes I, IIa, IIb and III according to Annex VIII. Under 4.1 Rule 1 it is defined:

"**All non-invasive devices are classified as class I, unless one of the rules set out hereinafter applies.**"

Rule 6.3 states:

"Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause:

- death or an irreversible deterioration of a person's state of health, in which case it is in class III; or
- a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb.
All other software is classified as class I."

With regard to the quality requirements that must be demonstrated for medical devices, see Annex I, 15.1:
Diagnostic devices and devices with a measuring function, shall be designed and manufactured in such a way as to **provide sufficient accuracy, precision and stability for their intended purpose, based on appropriate scientific and technical methods**. The limits of accuracy shall be indicated by the manufacturer.

15.2 is also likely to be of importance:

“**The measurements made by devices with a measuring function shall be expressed in legal units conforming to the provisions of Council Directive 80/181/EEC**.”

This directive contains the following table:

1.2.3. SI derived units with special names and symbols

Quantity	Unit		Expression	
	Name	Symbol	In terms of other SI units	In terms of SI base units
Plane angle	radian	rad		$m \cdot m^{-1}$
Solid angle	steradian	sr		$m^2 \cdot m^{-2}$
Frequency	hertz	Hz		s^{-1}
Force	newton	N		$m \cdot kg \cdot s^{-2}$
Pressure, stress	pascal	Pa	$N \cdot m^{-2}$	$m^{-1} \cdot kg \cdot s^{-2}$
Energy, work; quantity of heat	joule	J	$N \cdot m$	$m^2 \cdot kg \cdot s^{-2}$
Power ([1]), radiant flux	watt	W	$J \cdot s^{-1}$	$m^2 \cdot kg \cdot s^{-3}$
Quantity of electricity, electric charge	coulomb	C		$s \cdot A$
Electric potential, potential difference, electromotive force	volt	V	$W \cdot A^{-1}$	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-1}$
Electric resistance	ohm	Ω	$V \cdot A^{-1}$	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-2}$
Conductance	siemens	S	$A \cdot V^{-1}$	$m^{-2} \cdot kg^{-1} \cdot s^3 \cdot A^2$
Capacitance	farad	F	$C \cdot V^{-1}$	$m^{-2} \cdot kg^{-1} \cdot s^4 \cdot A^2$
Magnetic flux	weber	Wb	$V \cdot s$	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-1}$
Magnetic flux density	tesla	T	$Wb \cdot m^{-2}$	$kg \cdot s^{-2} \cdot A^{-1}$
Inductance	henry	H	$Wb \cdot A^{-1}$	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-2}$
Luminous flux	lumen	lm	$cd \cdot sr$	cd
Illuminance	lux	lx	$lm \cdot m^{-2}$	$m^{-2} \cdot cd$
Activity (of a radionuclide)	becquerel	Bq		s^{-1}
Absorbed dose, specific energy imparted, kerma, absorbed dose index	gray	Gy	$J \cdot kg^{-1}$	$m^2 \cdot s^{-2}$
Dose equivalent	sievert	Sv	$J \cdot kg^{-1}$	$m^2 \cdot s^{-2}$
Catalytic activity	katal	kat		$mol \cdot s^{-1}$

([1]) Special names for the unit of power: the name volt–ampere (symbol "VA") when it is used to express the apparent power of alternating electric current, and var (symbol "var") when it is used to express reactive electric power. The "var" is not included in GCPM resolutions.

Article 52 (6) states:

“Manufacturers of class IIa devices, other than custom-made or investigational devices, shall be subject to a conformity assessment as specified in Chapters I and III of Annex IX, and including an assessment of the technical documentation as specified in Section 4 of that Annex of at least one representative device for each category of devices.”

In further details extensive quality assurance measures, which also include regular reports based on a monitoring plan are specified.

In summary, the following points can be considered particularly relevant:

- With the instructions for use (technical manual) the manufacturer makes a decisive contribution to the classification as a medical device.
- Software that automatically generates measurement results that automatically contribute to diagnosis falls into the category MP IIa.
- Quality must be assured by means of a comprehensive process. The technical focus is on accuracy, precision and stability. In this context, the reference to the units of measurement to be taken into account is also likely to be relevant.

This raises the question of whether psychometric tests, which contribute to diagnosis by using software, should be assessed as medical devices and then consequently as medical devices in category IIa. In order to discuss this, the following section will first outline once again what is understood in psychology by a psychometric test based on measurement theory.

What is a psychometric test?

The EFPA Board of Assessment defines a psychometric test in its test evaluation guidelines (EFPA Board of Assessment, 2013):

“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."

The crucial aspect here is the aspect of measurement according to a standardized process. This measurement is done by assigning numbers to observations according to certain rules (Stevens, 1946). It is important to emphasize that this assignment of numbers does not yet yield an interpretable quantity, since psychometric measurements do not have a natural unit (Michell, 2001). Rather, the interpretation of the measurement is made by referring to a comparative value (e.g., by converting it to a norm value such as IQ in norm-oriented testing or comparing it to a cut-off value derived from a comparison sample in criterion-oriented testing). Thus, these interpretations do not have a unit of measurement that is one of those listed in Directive 80/181/EEC.

Tests based on measurement theory do provide information which is taken into account in clinical-psychological diagnostics. However, this is only done after interpretation by diagnosticians and consideration of further information (Diagnostics and Testing Board of the Federation of German Psychological Associations, 2017; Witteman, Van der Heijden, & Claes, 2018).

Several national (DTK, 2018; NIP in Evers et al 2010; COP in Muñiz et al 2011; BPS in Lindley & Bartram 2012) and international guidelines (AERA, APA, & NMCE, 2014; EFPA Board of Assessment, 2013) exist for assessing the quality of psychometric tests. All of these guidelines have in common that they define quality based on various criteria. These include reliability (measurement accuracy, precision, stability) as well as objectivity, validity, norm group, fairness, etc. This means that the quality assessment is much broader than that provided for in Regulation (EU) 2017/745 for medical devices. The level of detail is also much greater in the guidelines. Thus, as a rule, there are precise checklists with the information to be checked in relation to the quality criteria.

With regard to Regulation (EU) 2017/745, the implementation of psychometric tests in software is relevant. It should be noted that, as a rule, there is no functional difference between a test performed by means of paper and pencil and a computer-based test with regard to the relevant passage of Regulation (EU) 2017/745. In both cases, the measurement is done by assigning numbers. Software-based this is done computer-based instead of using evaluation templates. The conversion into standard values is also computer-based, but using the same rules as applied by human test administrators. Thus, the error rate can be minimized by software-based testing.

However, the resulting report, i.e., the information, is the same. Therefore, the use of software has no particular diagnostic significance. It would be absurd to classify one and the same instrument once as a medical device (computer-based procedure) and once not ("manual" procedure).

Is a psychometric test a medical device?

Based on the information provided previously, it is concluded here that psychometric tests are not medical devices. This is also true in the case that the procedures are used in a software-based manner.

The decisive factor for this evaluation is that psychometric tests can be used in a software-based manner, but they do not constitute software in their own right. Otherwise, every digitally administered patient questionnaire, which can also request information that is taken into account in the diagnosis (e.g., age, gender, drug consumption), would be a medical device of category IIa with the corresponding necessary quality assurance measures. A word processing program with which a finding is created or a medical record is filled diagnostically also contributes to the creation of a diagnosis, but not in the sense of a technical measurement whose procedural quality would have to be assured. In the interpretation of the regulation, a legally relevant contribution of software should therefore be independent as well as substantial and rather technical in nature.

Furthermore, it is stated that the focus of quality assurance for medical devices is on measurement accuracy. This is justified, since measurements according to Directive 80/181/EEC have natural units, which are directly interpretable. This is not the case for psychometric tests, and the scores used therefore do not appear in Directive 80/181/EEC. An interpretation is made by comparison with an empirically obtained reference sample. This also applies to neuropsychological procedures that use reaction times, for example. For this reason, demonstrating the validity of this test score interpretation is of paramount importance for psychometric tests. Accordingly, the above guidelines for assessing the quality of psychometric tests place much emphasis on testing validity: does the procedure measure what it claims to measure? This difference between measurement in the medical versus psychological sense is of considerable importance and supports the statement not to classify tests based on measurement theory as medical devices.

Recommendation and consequences

It is clearly recommended that psychometric tests should not be classified as medical devices, even if they are administered using software.
From the EFPA's point of view, it should be emphasized that due to the differences in measurement quality assurance for psychometric tests has a much broader focus.

Classification as a medical device could lead to a perspective decline in the quality of psychometric tests due to a focus on measurement accuracy and a neglect of validity evidence. This would have potentially dramatic implications for the quality of diagnoses. In the opinion of the EFPA, the quality of psychometric tests should be assessed by psychologists based on relevant standards published by professional psychological organizations.

Literature

AERA, APA, & NMCE. (2014). *Standards for Educational & Psychological Testing*. Washington, DC, USA.

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2017). Qualitätsstandards für Psychologische Gutachten. In.

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2018). Testbeurteilungssystem des Diagnostik und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. *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.

Lindley, P. A., & Bartram, D. (2012). Use of the EFPA Review Model by the UK and Issues Relating to the Internationalization of Test Standards. *International Journal of Testing,* 12, 108-121.

Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist, 36*, 211-218.

Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, Á., & Peña-Suárez, E. (2011). Evaluación de tests editados en España [Assessment of tests published in Spain]. *Papeles del Psicólogo,* 32(2), 113–128.

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.