ROCZNIKI PSYCHOLOGICZNE / ANNALS OF PSYCHOLOGY
2016, XIX, 3, 453-468
DOI: http://dx.doi.org/10.18290/rpsych.2016.19.3-2en

JERZY MARIAN BRZEZIŃSKI
Adam Mickiewicz University in Poznań
Institute of Psychology

ON THE METHODOLOGICAL PECULIARITIES OF SCIENTIFIC RESEARCH AND ASSESSMENT CONDUCTED BY CLINICAL PSYCHOLOGISTS

The article focuses on the methodological peculiarities of research practice (scientific research and assessment) in clinical psychology. The author indicates typical departures of this practice from the methodological standards of modern psychology – particularly from those that constitute the patterns of evidence-based assessment and evidence-based practice in psychology: ignoring the primary role of psychological theories, too frequent departures from psychometric standards in the procedures of constructing and using psychological tests (which are sometimes pseudotests, such as Koch's Tree Test). The importance of construct validity is stressed in the article.

Keywords: psychological assessment; psychological theory; research process; EBA; EBPP; internal validity; external validity; clinical psychology.

What induced me to address this topic was the re-reading, after about 30 years, of a classic article by Saul Rosenzweig (1907-2004),[1] one of the most original clinical psycyhologists. The article I have in mind is "The Experimental Situation as a Psychological Problem" (Rosenzweig, 1933). In his text, S. Rosenzweig wrote about the "peculiarities" of experimental research in psychology. In brief, he pointed out that:

– the researcher becomes an element of the research situation;

– the subject's behavior in the research situation is influenced by variables related to and characterizing the subject, such as personality, motivation, etc.;

Address for correspondence: JERZY MARIAN BRZEZIŃSKI – Institute of Psychology, Adam Mickiewicz University in Poznań, ul. Szamarzewskiego 89AB, 60-568 Poznań; e-mail: brzezuam@amu.edu.pl

[1] About Saul Rosenzweig – cf. Kaufman (2007).

– interaction is established between the researcher and the subject.

What I would like to consider in this methodological outline – going beyond the subject matter discussed by S. Rosenzweig and, simply, accepting it (cf. also: Brzeziński, 1994) – is the problem of the peculiarities in scientific research and assessment (as seen from the perspective of "hard" methodological standards) conducted by clinical psychologists (further referred to as clinicians). As regards the associations described by S. Rosenzweig, I would consider them as belonging to the *social psychology of psychological research*, as in the title of Arthur G. Miller's (1972) book, well-known in the 1970s.

Naturally, clinical psychology is, quite simply, a branch of psychology *sensu proprio* (understood as an empirical science and not otherwise). Let me stress right away that it is not a separate psychology. We have taken the following stand on this issue (Brzeziński & Toeplitz-Winiewska, 2008, p. 305; also: Brzeziński, 1996, 2013):

> In the "eternal" dispute between academic psychologists and practicing psychologists (for this is, more or less, where the demarcation line runs), our stance . . . is analogous to the one voiced by Matarazzo (1987) or Ellis (1992), and can be phrased as follows: there is only one psychology, not many applied psychologies; its empirical findings are applied in various fields of social practice – as a response of the domain of science to the social demands that are reported.

Consequently – and this is not only my point of view – the methodological "peculiarity" of research and assessment practice in clinical psychology is not connected with handling specific "orders" from the domain of social practice (in accordance with the schema I described in: Brzeziński, 2013). After all, in this respect (in other words: at this level of generality) it does not differ from such specialized and practically oriented branches of psychology *sensu proprio* as work psychology, penitentiary psychology, or educational psychology. Let us therefore stress, again, that the practically oriented "sub-psychologies" I have mentioned as examples are not distinct psychologies in the methodological sense. This, however, does not mean that clinicians' research practice (scientific and diagnostic) is not particularly susceptible to departures from the methodological standards of psychology *sensu proprio*. It is this orientation, characterized by disregard for empirical psychological theories and, as a result, by a drift towards para-science, that I would like to discuss. Of course, ignoring modern methodological standards, disregard for empirical psychological theories, or excessive attachment to clinicians' personal experience (practice) as the main source of assessment knowledge is not the only peculiarity of clinical psychology. It is, however, impossible to address all issues in a study as short as this one.

From scientific theory to empirical research, but not without theory

On various occasions, psychologists (particularly those who argue that the profession of psychologist should be practiced in accordance with serious psychological theory) have made it a habit to cite the words of famous scientists emphasizing the significance of theory to effective practice (e.g., clinical practice – assessment or psychotherapy). Words that have been especially popular among psychologists are those of physicist Robert Kirchhoff (quite often wrongly attributed to psychologist Kurt Lewin): "The most practical thing is a good theory." It is also possible to quote the words of sociologist David Silverman (2006, p. 14): "Without a theory, such phenomena as 'gender', 'personality', 'talk' or 'space' cannot be understood by social science. In this sense, without a theory there is nothing to research," or those of biologist François Jacob (1973, p. 15): "In the dialogue between theory and experience, theory always has the first word. It determines the form of the question and thus sets limits to the answer."

However, before I highlight the "theory–experience (practice)" relationship with reference to clinicians' practice, let me first present the methodological standard regulating the contemporary research procedure in psychology (more broadly: in social sciences). It is presented in Fig. 1, which shows the theoretical components of the research procedure (at least as I see them).

To begin with, research practice is "immersed" in a theoretical context. Outside that context, any research or assessment makes no sense at all – perhaps this sounds too strong, but such is my opinion; this senselessness manifests itself with particular clarity when clinicians get down to "improving" their clients' psyche, and all they have to offer is their personal "experience" (what, then, would they need the label of "psychologist" for?). Still, there is a problem with a psychological theory, methodologically correct and rich in meaning. Unfortunately, we do not have many of those in psychology. But I would like to believe that, with an increase in *methodological awareness* among psychologists (largely thanks to psychological studies conducted at good *research universities*, i.e., research-oriented ones – an American idea that I would like to see transplanted to Poland), true theories will drive out their attractively wrapped imitations in the form of pulp psychological literature.

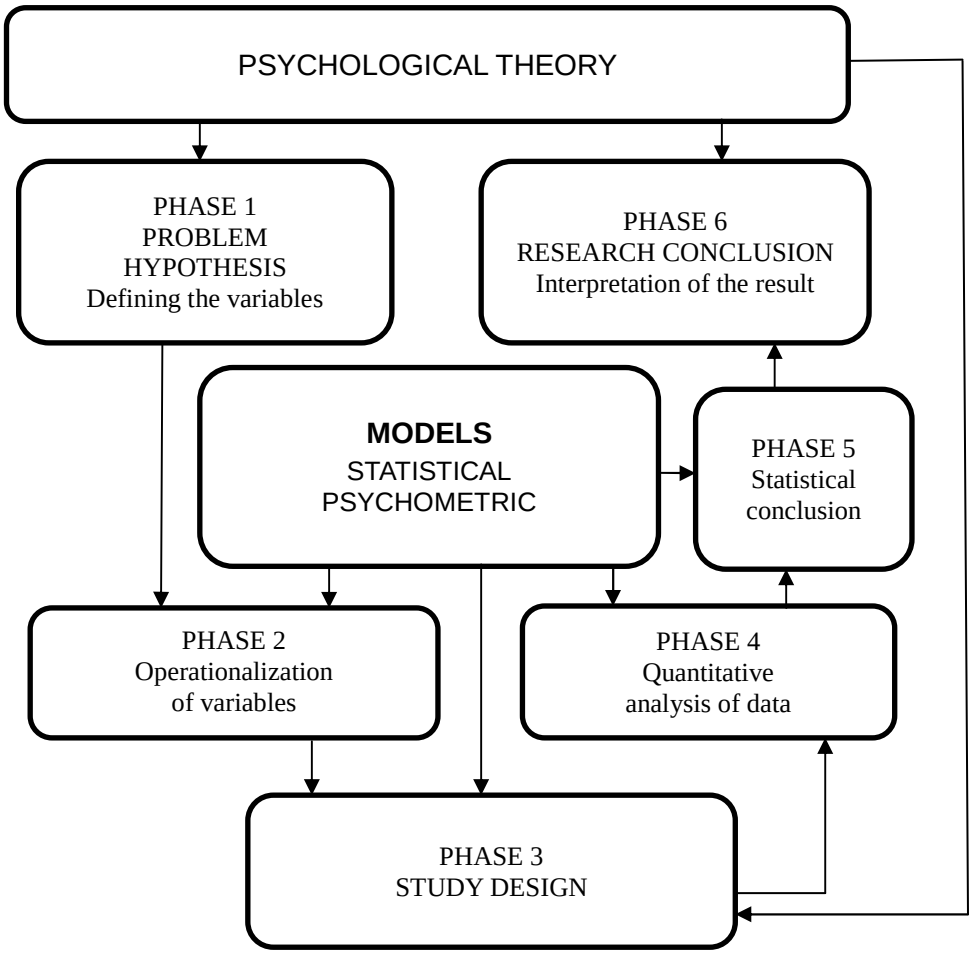

Figure 1. Research procedure.

Theory provides the researcher with language – with a "grammar" and a "lexicon" – in which it is possible to build (psychologically) meaningful sentences. These in turn not only serve to describe and explain phenomena, but also act as theoretical support for the professional assistance provided by clinicians. Science is multiparadigmatic (and so is psychology). The researcher makes a choice (which, optimally, should be fully conscious) between *paradigms* (as understood by Kuhn, 1996). Choosing one of the paradigms enables him or her to "descend" to a lower level of theoretical analysis and empirical analysis. But it

is easy to fall into trouble when trying to combine elements of theory derived from different, mutually exclusive, paradigms. All theoretical changes introduced into a theory must be defined in the language of the same paradigm. The researcher should follow the basic principle of *paradigmatic noncontradiction*. Clinicians, who usually conduct research at the subparadigmatic level, do not always realize the importance of this principle. It is in the language of a particular theory or noncontradictory theories with a unifying paradigm that the dependent variable is defined – as well as independent ones constituting *the picture of variables significant to a given dependent variable* $O(P_Y)$. Based on variables thus defined, the researcher formulates problems and hypotheses (Phase 1).

Every psychological theory requires empirical interpretation. Theoretical terms (variables) have to be linked with observational terms. In other words, transition to Phase 2 requires giving empirical meaning to theoretical variables. What is therefore very important is the operationalization of the dimensions of researcher-constructed $O(P_Y)$, consistent with the theory. The program of the operationalization of variables significant to Y and of Y itself must be derived from a theory and – in particular – noncontradictory to the theory. A program of the operationalization of variables is therefore methodologically incorrect if it refers to different theories (more precisely, to definitions of theoretical variables based on theories not compatible with one another – belonging to different paradigms). For how can Hermann Rorschach's method (what empirical psychological theory gives meaning to it?) be combined with Aaron Beck's Depression Inventory? It is obvious, too, that the language of the interpretation of measurement results is the language of the same theory in which variables have been defined and based on which the research hypotheses have been formulated (or one that is noncontradictory to it). Consequently, the researcher is obliged to follow *the principle of consistency of the operationalization program with the theoretical program.*

The psychological tests or countless personality questionnaires used, quite eagerly, by clinicians require a few words of comment (I exclude the so-called projective methods, popular in the clinicians' community: *the Rorschach method*, *the Thematic Apperception Test*, or the *Rotter Incomplete Sentence Blank*, as having dubious scientific value). The method especially popular in that community is *David Wechsler's Intelligence Scale*, and it is since the construction of its first version – the *Wechsler-Bellevue Intelligence Scale (W-B*) by D. Wechsler in 1939 that clinicians have been using it not only as a measure of IQ, but also as a kind of test of personality and clinical deviations from the norm (cf. Wechsler,

1941/1998; Rapaport, 1945; Zimmerman & Woo-Sam, 1973; Frank, 1984; Kowalik, 1998; Tulsky, Saklofske, Chelune et al., 2003).

The Wechsler Intelligence Scales are an example of a method that was constructed on the basis of the diagnostic experience and intuitions of its creator, who collected the most useful psychological assessment tests "under one roof" to make up a kind of test battery (Boake, 2002). In the "theoretical" sense, it was united by his definition of intelligence as:

> . . . the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is aggregate because it is composed of elements or abilities which though not independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence (Wechsler, 1939/1998, p. 16).

What psychologists attempted to do later was reconstruct the theoretical basis of this test method, reputedly the most popular one in the world. As stated in the manual for the 1997 *Wechsler Adult Intelligence Scale* (WAIS-III):

> The development of Wechsler's tests was not based on theory (except perhaps on Spearman's [1927] *g*, or general intelligence theory) but instead on practical and clinical perspectives. . . . Wechsler's view of IQ tests was that they were a way to peer into an individual's personality. Years after the development of the Wechsler scales, extensive theoretical speculations have been made about the nature and meaning of these tests and their scores, but originally the tests were developed without regard to theory (Kaufman & Lichtenberger, 1999, p. 3).

The above quotation points to the lack of original theoretical foundations not only for this version of the *Intelligence Scale*, but also for the previous versions (cf. Hornowska, 1988). In the years that have elapsed since the publication of the *Wechsler-Bellevue Intelligence Scale* attempts have been made to perform an ex post reconstruction of the theoretical basis of Wechsler's scales. Scholars have tried to "locate" tests making up the Wechsler battery in Charles Spearman's *g factor theory* (measuring the saturation of each test with the *g* factor – cf. Leckliter, Matarazzo, & Silverstein, 1986). The two important intelligence theories that have been seen as linked with Wechsler's tests are: Raymond B. Cattell's and John L. Horn's *theory of fluid and crystallized intelligence* and Joy P. Guilford's *structure of intellect theory*. These attempts to find the theoretical basis of Wechsler's *Intelligence Scale* are also mentioned in the WAIS-III and WAIS-IV manuals (cf. Kaufman & Lichtenberger, 1999; Lichtenberger & Kaufman, 2009).

As regards personality questionnaires, of the three questionnaire construction strategies listed by Bogdan Zawadzki (2006, p. 77): theoretical (also called deductive), external (also referred to as criterion), and internal (also called induc-

tive), it is the first one that is definitely consistent with the basic assumption that theory precedes the operationalization of variables. Today, after the publication of the groundbreaking psychometric study that led to the recognition of one more, crucial type of validity – *construct validity* (Cronbach & Meehl, 1955; see also Cronbach, 1989), there remains no doubt that outside empirical psychological theory a test is merely a worthless imitation. Commenting on the classic text by Lee J. Cronbach and Paul E. Meehl, Drew Westen and Robert Rosenthal (2003) wrote significant words about construct validity and its place in contemporary psychological thought – words that one can hardly disagree with:

> Construct validity is one of the most important concepts in all of psychology. It is at the heart of any study in which researchers use a measure as an index of a variable that is not itself directly observable (e.g., intelligence, aggression, working memory). If a psychological test (or, more broadly, a psychological procedure, including an experimental manipulation) lacks construct validity, results obtained using this test or procedure will be difficult to interpret (p. 608).

Similarly, the authors of the extension of the *classical test theory* believe that, as they wrote in their fundamental study (cf. Lord & Novick, 1968), construct validity is "the most important characteristic of a test" (p. 278). This is undoubtedly the most important aspect of determining the validity of a test.

In brief, a test that is not based on an empirical theory is not worthy of that name. One of the well-established ways of determining construct validity is the procedure of *convergent and discriminant validation by the multitrait–multimethod matrix*, proposed by Donald T. Campbell and Donald W. Fiske (Campbell & Fiske, 1959).

Fig. 1 shows that, apart from empirical psychological theory, which determines the construct validity of a test, there are two other, nonpsychological theories (models) – *statistical* (defining the framework of quantitative interpretation of the test result) and *psychometric*, in which the test is constructed and based on which the main parameters of the test's *psychometric goodness* are established – above all, its reliability and the size of the *standard error of measurement*. They make it possible to build a *confidence interval* for the *true result* (in accordance with the *classical test theory* by Harold O. Gulliksen, 1950). Knowing it is essential for the correct interpretation of the test result (cf. AERA/APA/NCME, 1999/2007). Unfortunately, the knowledge of statistical and psychometric standards is not clinicians' forte. Many of them are confident about their clinical supercompetence and reject the statistical approach – as though Paul E. Meehl's famous book debunking such naive beliefs, *Clinical Versus Statistical Prediction.*

A Theoretical Analysis and a Review of the Evidence (Meehl, 1954), had not been published many years ago.

In Phase 3, the empirical study design is prepared. What is necessary here is familiarity with the statistical background of experimental and correlational research – e.g., ANOVA/MANOVA model for experimental studies (with reference to the principle of randomization). And again, without the knowledge of psychometric and statistical models it will be impossible to implement Phases 4 and 5 correctly.

In Phase 6 of the study, the interpreted result (statistical conclusion) of the statistical method applied (measures of correlation, significance of differences tests) should be transformed into a research conclusion. I am writing about this because the researcher is too often content with stating that his or her hypothesis regarding the relationship between variables or regarding the significance of the difference between the mean values of the dependent variable in the experimental and control groups has been confirmed. A research "success" is to obtain a significance level of $p = .05$. It is good if the information about the value of p is supplemented with information about the value of *effect size* (ES; e.g., Cohen's d coefficient applied to the results of Student's t-test) – about the real strength of the impact of variable X on variable Y (for several years, this has been the standard requirement for authors of empirical texts; cf. APA, 2010; see also Wilkinson, 1999). The values of ES should be interpreted in terms of intervals (JARS Group, 2008; Wilkinson, 1999).

I do not equate the *statistical conclusion* (the sequence indicated above: the value of test statistic, e.g., "Student's t value → statistical significance level, e.g., not exceeding $p = .05$ → ES value, e.g., Cohen's d") with the *research conclusion*. I subscribe to the position articulated by Bruce M. King and Edward W. Minium (King & Minium, 2009, p. 25) regarding the division of conclusions into *statistical conclusions* and *research conclusions*. I would also add two justifications for this distinction. The first one is the evaluation of whether or not the conditions in which a given statistical test was applied affected its result. After all, the compared samples are not, in fact, selected by sampling with replacement from the population (and such is the assumption of the statistical significance test model). Did the researcher really control all external influences that may have distorted the participants' behavior in the study situation? Of course not. Did the researcher apply the randomization principle in experimental research? It is therefore necessary to return to Phase 3 and critically re-examine the study design. For example: when there is justified suspicion that there may be a *pretest effect* in an experimental study, the study should be conducted according to the *Solo-*

mon design. The second justification is the assessment of the practical consequences of the risk, accepted by the researcher, of making an I^0 error (always in the phase of planning the study, not after its completion!). An excessively strict *p* level may result in "sinking" an interesting hypothesis, while an excessively liberal *p* level may lead to the dissemination of a false result, which will become the basis for, say, a therapeutic procedure that may turn out to be harmful (e.g., as when an ill-tested medicine proves to be toxic). Perhaps, then, it is not advisable to adhere rigidly to the level of $p = .05$? Perhaps it is sometimes necessary to apply the level of $p = .001$, and sometimes $p = .10$ is enough (e.g., in exploratory studies)?

Also in this case – particularly in the context of studies conducted by clinicians in field conditions (e.g., in clinics) rather than in strictly controlled laboratory conditions – it is necessary to stress the importance of *external validity* (especially!) and *internal validity* of the study design. On the one hand, to ensure high internal validity, it would be advisable – as much as possible – to isolate the study from confounding external influences and conduct it in laboratory conditions, in accordance with the *principle of randomization* (maximizing internal validity). On the other, such isolation, sometimes brought to perfection, makes it unreal and artificial, with external validity considerably lowered. Leaving the laboratory decreases the internal validity of the study but increases its external validity. This is one of the basic problems of research on psychotherapy following two standards: *efficacy vs. effectiveness* (cf. e.g., Hunsley, Elliot, & Therrien, 2013).

Unfortunately, the intention to meet high methodological standards in order to ensure that a clinical study has high internal validity is usually not accompanied by ensuring its high external validity. *Internal validity* requires high precision of measurement, control groups, randomization, isolation from external influences, etc. The internal validity criteria are met by *RCT* studies (*Randomized Controlled Trials*). As regards *external validity*, it requires making the study conditions close to real-life conditions and conducting the study on representative groups. An example could be Martin Seligman's study on the effectiveness of psychotherapy, conducted using *Consumer Reports* (Seligman, 1996; for a critique of that study, see: Nathan, Stuart, & Dolan, 2000; Jaworska, 2001). What is the good of a study conducted on a *representative sample* (as Seligman stressed in his defense) as a postal survey, having relatively high *external validity*, if its internal validity was unacceptable (e.g., no control group).

To sum up this "validity" thread of my paper, I would like to draw special attention to the sources of the low validity of studies conducted by clinicians:

internal validity:
lack of a comparison (control) group
lack of randomization
external validity:
low representativeness of the sample
low representativeness of the study conditions

Is it possible to prioritize one type of validity over another – *external validity* over *internal validity* – in order to obtain a result that is closer to the expectations, as clinicians seem to do? Let me give a methodologist's answer. No, it is not, for what is the good of *external validity* in the case of research results that are rather dubious when it comes to their *internal validity*. However, I do realize – also as a methodologist – that it is impossible to fully meet both of these conditions simultaneously. This can only happen in an idealized world, which the one surrounding us is not. All the researcher can do is try to attain some kind of compromise. It is also possible to sacrifice one type of validity "a little" for the sake of the other.

Summing up, empirical psychological theory and the models (theories) concomitant with it – statistical and psychometric – determine the shape of scientific research:

– as the source of definitions of variables,

– as the basis of the operationalization of variables (giving them empirical meaning),

– as the framework for the quantitative (statistical) interpretation of data,

– as the framework for the the psychological interpretation of the research result (treated as aggregate empirical data).

From empirical psychological theory to EBA and EBPP

Since the 1990s, following the example of medicine,[2] emphasis has been placed on the need for psychologists' assessment to be based on empirical evidence (*evidence-based assessment, EBA* – cf. Stemplewska-Żakowicz & Paluchowski, 2008; Stemplewska-Żakowicz, 2009; Paluchowski, 2010). This model of assessment procedure is part of the more general model of *evidence-based practice*

[2] In April 2015, I took part in a debate (conducted in accordance with the Oxford Union Debate formula: "This house believes that . . .") organized by the editing board of *Nowotwory. Journal of Oncology.* Its participants were instructed to present their positions ". . . in the spirit of evidence-based medicine, i.e., on the basis of reliable and up-to-date scientific research . . ."

in psychology (EBPP), which American Psychological Association (cf. APA, 2006) defined as follows: "Evidence-based practice in psychology (EBPP) is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (p. 273).

In fact, in Polish clinical psychology, formed to a great extent by Andrzej Lewicki's works (in the 1960s – cf. Lewicki, 1969), assessment has been modeled for a long time on the research process – that is, with a dominant role of theory and reliable methodology (cf. e.g., Brzeziński & Kowalik, 1991; Kowalik & Brzeziński, 1991a, 1991b). Years ago, the founder of Poland's first Department of Clinical Psychology at the Adam Mickiewicz University in Poznań, A. Lewicki (1969, p. 84), wrote:

> From the methodological point of view, **the clinical psychologist's assessment activity should be regarded as a form of scientific research**, applied to solving practical clinical problems. The essence of every scientific study is: (a) to formulate a problem, (b) to put forward a hypothesis, i.e., a probable solution to the problem, and (c) to test the hypothesis by applying appropriate research methods. Clinical research meets these conditions. [my emphasis]

Is this not a vision of EBA? I do agree with the EBA model when it comes to the standard of assessment practice, and especially when it comes to eliminating "shamanic" assessment practices, relying on psychological pseudotests such as *the Szondi Test*, *the Lüscher's Color Test*, or *Koch's Tree Test*, from psychologists' repertoire of instruments (and especially from the repertoire of those clinicians who have become enchantedly attached to them). I subscribe to the thorough criticism of such "tests" offered by Katarzyna Stemplewska-Żakowicz (2009). It must be noted that these three "tests" do not exhaust the black list of scientifically illegitimate assessment instruments fondly used by a considerable part of clinicians' community. An important supplement to this list could be another black list of procedures by means of which incompetent para-clinicians or ones seeking easy money (witness the various "psychologists' offices") influence their naive clients; the list would include Bert Hellinger's family constellations, NLP, or various magic psychoanalytic practices. I do not think they meet the conditions of EBPP.

Characterizing one of the important peculiarities of the EBA model, Władysław J. Paluchowski (2010, p. 11) writes that this model rejects the assumption that

> . . . observations derived from clinical practice (one's own or other people's) constitute a reliable and sufficient source of practical medical knowledge. The value of personal experience and practice is overestimated, since they always have a local character, limited to cases

> not representative of all patients but of contextually selected ones (biased sample), and their evaluation depends on the physician's specific personal characteristics (the type of education or cognitive preference). Moreover, reliance on the experience of authorities is more often based on faith than on wisdom, which sometimes gives way to established stereotypes and fashion (Sułkowski, 2007).

In my opinion, this also refers to clinicians' traditional practice.

However, I would like the role of psychological empirical theory to be stressed more clearly, since it is the factor that determines the value of psychological (also clinical) assessment above all. Proper emphasis was given to theory in the co-called Daubert guidelines prepared by the U.S. Supreme Court, which are used with regard to expert opinions in the American judiciary.[3] The seven Daubert guidelines (as cited in Ritzler, Erard, & Pettigrew, 2002, pp. 202-203) are as follows:

> (1) Is the proposed **theory** (or technique), on which the testimony is to be based, testable?
>
> (2) Has the proposed **theory** (or technique) been tested using valid and reliable procedures and with positive results?
>
> (3) Has the **theory** (or technique) been subjected to *peer review*?
>
> (4) What is the known or potential error rate of the **scientific theory** or technique?
>
> (5) What standards, controlling the technique's operation, maximize its validity?
>
> (6) Has the **theory** (or technique) been generally accepted as valid by a relevant scientific community? (Grove & Barden, 1999, p. 226)
>
> (7) [Added later] Do the expert's conclusions reasonably follow from applying the **theory** (or technique) to this case? (Grove & Barden, 1999, p. 226, my emphasis)

Let us note that the most frequently recurring term is "theory." If clinicians preparing expert opinions (e.g., for courts: deciding on the degree of mental capacity, sexology, divorces, etc.) fail to keep up with the technological and methodological development of psychology (theories and instruments), they condemn themselves to humiliation (e.g., by hanging on to Koch's *Tree Test*) as well as to exclusion from serious debates and professional orders (though not necessarily from easy money).

[3] Jason Daubert and Eric Schuler filed a lawsuit against Merrell Dow Pharmaceuticals Inc., a pharmaceutical concern, because they believed they had been born with physical defects as a result of their mothers taking Bendectin medicine during pregnancy. The court decided that the expert opinions prepared in the course of the proceedings had to meet what was called the Daubert guidelines. These guidelines became the recommendations for judges in the American judiciary.

Conclusion

The research practice of clinical psychologists (both scientific research and assessment) is not yet fully based on the methodological standards of empirical psychology, in which there is no place for speculations, self-proclaimed as "theories" but not having anything to do with science – frequently of psychoanalytic provenance (and psychoanalysis, as K. Popper demonstrated, diverges from the rational model of science by not being *falsifiable*).

What largely determines the scientific character of a clinical scientific (or assessment) study – as I have tried to show when discussing Fig. 1, emphasizing the role of empirical psychological theory, psychometric models, and statistical models – is reliance on an empirically tested psychological theory.

Assessment practice today refers to EBA standards; it also refers (see Daubert criteria) to proven theories and correctly constructed instruments, providing *replicable* results (this, too, is a fairly effective barrier against scientific fraud). There is no place among them for a belief in having special competence due to the years of practice or for personal feelings and impressions from contact with the subjects as sources of reliable empirical data (cf. the list of *Multiple Types of Research Evidence* included in the report of APA Presidential Task Force on Evidence-Based Practice, 2006, p. 274).

Improving clinicians' research practice means developing reliable techniques and obtaining replicable data.

REFERENCES

AERA/APA/NCME; American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999/2007). *Standardy dla testów stosowanych w psychologii i pedagogice*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

APA, American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

APA, APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271-285.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*(3), 383-405.

Brzeziński, J. (1994). Dimensions of diagnostic space. In J. Brzeziński (Ed.), *Probability in theory-building. Experimental and non-experimental approaches to scientific research in psychology* (pp. 197-223). Amsterdam-Atlanta, GA: Rodopi. Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 39.

Brzeziński J. (1996). Theory and social practice. One or two psychologies? In A. Zeidler-Janiszewska (Ed.), *Epistemology and history. Humanities as a philosophical problem and Je-*

rzy Kmita's approach to it (pp. 351-364). Amsterdam-Atlanta, GA: Rodopi. Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 43.

Brzeziński, J. (2013). Methodological awareness and ethical awareness in the context of university education (on the example of psychology). In B. Bokus (Ed.), *Responsibility. A cross--disciplinary perspective* (pp. 261-277). Warsaw, PL: Lexem.

Brzeziński, J., & Kowalik, S. (1991). Diagnoza kliniczna w kontekście praktyki społecznej. In H. Sęk (Ed.), *Społeczna psychologia kliniczna* (pp. 187-212). Warsaw, PL: Wydawnictwo Naukowe PWN.

Brzeziński, J., & Toeplitz-Winiewska, M. (2008). Model zawodowy psychologa klinicznego. In H. Sęk (Ed.), *Psychologia kliniczna* (2nd ed., vol. 1, pp. 299-324). Warsaw, PL: Wydawnictwo Naukowe PWN.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*(2), 81-105; also Polish translation in J. Brzeziński (Ed.). (2005). *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (pp. 431-459). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Cronbach, L. J. (1989). Construct validation after 30 years. In R. L. Linn (Ed.), *Intelligence. Measurement theory and public policy* (pp. 147-171). Champaign: University of Illinois Press.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302; also Polish translation in J. Brzeziński (Ed.) (2005). *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (pp. 403-430). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Ellis, H. C. (1992). Graduate education in psychology. Past, present, and future. *American Psychologist, 47*(4), 570-576.

Frank, G. (1984). *Wechsler enterprise*. Oxford: Pergamon Press.

Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, and Law, 5*(1), 224-242.

Gulliksen, H. O. (1950). *Theory of mental tests.* New York: J. Wiley.

Hornowska, E. (1988). Skala inteligencji Davida Wechslera w świetle teorii inteligencji. In J. Brzeziński & E. Hornowska (Eds.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (2nd ed., pp. 21-48). Warsaw, PL: Państwowe Wydawnictwo Naukowe.

Hunsley, J., Elliot, K., & Therrien, Z. (2013). The efficacy and effectiveness of psychological treatments. Retrieved May 3, 2015, from http://www.cpa.ca/docs/File/Practice/TheEfficacyAndEffectivenessOfPsychologicalTreatments_web.pdf

Jacob, F. (1973). *The logic of life: A history of heredity*. New York: Pantheon Books.

JARS Group, APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology. Why do we need them? What might they be? *American Psychologist, 63*(9), 839-851.

Jaworska, A. (2001). Badanie efektywności psychoterapii metodą Consumer Reports: Analiza metodologii. *Czasopismo Psychologiczne, 7*(2), 227-240.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III Assessment*. New York: Wiley.

Kaufman, M. (2007). In memoriam: The idioverse of Saul Rosenzweig (1907–2004). *Journal of Psychotherapy Integration, 17*(4), 363-368.

King, B. M., & Minium, E. W. (2009). *Statystyka w psychologii i pedagogice*. Warsaw, PL: Wydawnictwo Naukowe PWN.

Kowalik, S. (1988). Diagnostyczne i kliniczne właściwości Skal Inteligencji: *W-B I/II, WAIS, WAIS-R*. In J. Brzeziński & E. Hornowska (Eds.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (2nd ed, pp. 386-462). Warsaw, PL: Wydawnictwo Naukowe PWN.

Kowalik, S., & Brzeziński, J. (1991a). Diagnoza kliniczna. In H. Sęk (Ed.), *Społeczna psychologia kliniczna* (pp. 213-239). Warsaw, PL: Wydawnictwo Naukowe PWN.

Kowalik, S., & Brzeziński, J. (1991b). Protodiagnoza kliniczna. In H. Sęk (Ed.), *Społeczna psychologia kliniczna* (pp. 240-254). Warsaw, PL: Wydawnictwo Naukowe PWN.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: The Chicago of University Press.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology, 42*, 322-342.

Lewicki, A. (1969). Psychologia kliniczna w zarysie. In A. Lewicki (Ed.), *Psychologia kliniczna* (pp. 9-155). Warsaw, PL: Państwowe Wydawnictwo Naukowe.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. New York: Wiley.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Matarazzo, J. D. (1987). There is only one psychology, no specialties, but many applications. *American Psychologist, 42*, 893-903.

Meehl, P. E. (1954). *Clinical versus statistical prediction. A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Miller, A. G. (Ed.). (1972). *The social psychology of psychological research*. New York: The Free Press.

Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin, 126*(6), 964-981.

Paluchowski, W. J. (2010). Diagnoza oparta na dowodach empirycznych – czy potrzebny jest „polski Buros"? *Roczniki Psychologiczne, 13*(2), 7-27.

Rapaport, D. A. (1945). *Diagnostic psychological testing* (vol. 1). Chicago, IL: Yearbook Publishing.

Ritzler, B., Erard, R., & Pettigrew, G. (2002). Protecting the integrity of Rorschach expert witnesses. A reply to Grove and Barden (1999) Re: The admissibility of testimony under *Daubert/Kumho* analyses. *Psychology, Public Policy, and Law, 8*(2), 201-215.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review, 40*(4), 337-354.

Seligman, M. E. P. (1996). Skuteczność psychoterapii. Badania ankietowe. *Czasopismo Psychologiczne, 2*(2), 101-112.

Silverman, D. (2006). *Interpreting qualitative data: Methods for analyzing talk, text and interaction*. (3rd ed.) London: SAGE Publications Ltd.

Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Stemplewska-Żakowicz, K. (2009). *Diagnoza psychologiczna. Diagnozowanie jako kompetencja profesjonalna*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Stemplewska-Żakowicz, K., & Paluchowski, W. J. (2008). Podstawy diagnozy psychologicznej. In J. Strelau & D. Doliński (Eds.), *Psychologia. Podręcznik akademicki* (vol. 2, pp. 23-63). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Sułkowski, H. (2007). *Co lekarz praktyk o medycynie opartej na dowodach wiedzieć powinien*. Retrieved May 3, 2015, from http://www.umed.lodz.pl/kmz/doc/ebm.pdf

Tulsky, D. S., Saklofske, D. H., Chelune, G. J., Heaton, R. K., Ivnik, R. J., Bornstein, R., Prifitera, A., & Ledbetter, M. F. (Eds.) (2003). *Clinical interpretation of the WAIS-III and WMS-III.* San Diego, CA: Academic Press.

Wechsler, D. (1939/1998). Definicja i natura inteligencji. In J. Brzeziński & E. Hornowska (Eds.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (2nd ed., pp. 15-20). Warsaw, PL: Wydawnictwo Naukowe PWN.

Wechsler, D. (1941/1998). Diagnostyczne i kliniczne właściwości Skali Wechsler-Bellevue. In J. Brzeziński & E. Hornowska (Eds.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (2nd ed., pp. 363-385). Warsaw, PL: Wydawnictwo Naukowe PWN.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84*(3), 608-618.

Wilkinson, L., & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

Zawadzki, B. (2006). *Kwestionariusze osobowości. Strategia i procedura konstruowania*. Warsaw, PL: Wydawnictwo Naukowe Scholar.

Zimmerman, I. L., & Woo-Sam, J. M. (1973). *Clinical interpretation of the Wechsler Adult Intelligence Scale*. New York: Grune & Stratton.