



## THE RELIABILITY OF PROJECTIVE TECHNIQUES AS TOOLS OF PSYCHOLOGICAL ASSESSMENT. PART 2: THE TREE TEST\*

Katarzyna STEMPEWSKA-ŻAKOWICZ<sup>1</sup>, Władysław Jacek PALUCHOWSKI<sup>2</sup>

<sup>1</sup> Faculty of Diagnostic Methods and Applied Psychology, Chodkowska University, Warszawa, Poland

<sup>2</sup> Institute of Psychology, Adam Mickiewicz University, Poznań, Poland

### Abstract

This is the second part of a two-part article on the reliability of projective techniques in psychological diagnosis for the purposes of the justice system. The authors advance the thesis that the conclusions produced by projective techniques, which were originally created and used in the clinical context, are not reliable enough to support court decisions. Numerous arguments in support of this thesis and examples of misapplied tools are presented in both parts of the article. The second part gives a description, historical account and detailed discussion of the Tree Test, which was developed by the Swiss psychologist, Karol Koch, in the 1940s. The Polish version of the textbook is a translation with no changes from the German language text book of 1947. The resulting anachronisms and the lack of a Polish adaptation are serious drawbacks of the tool. An extensive review of the literature demonstrates that there is a lack of evidence of the validity of the mode of interpretation of tree drawings described in the handbook; moreover the reliability of the measurements made using this technique is unknown. The analyst is encouraged to adopt practices that can be classified as the sort of flawed “clinical prediction” that Paul Meehl warned of. The authors state in conclusion that the tree test does not meet contemporary criteria for reliable diagnostic tools and could be harmful if used as such. As we know from survey evidence, however, the tree test and other controversial techniques are often used by expert psychologists in the opinions they prepare for the justice system. The article concludes with a discussion of the criteria of reliability for diagnostic tools used in psychology.

### Key words

Projective techniques; Personality assessment; Koch’s Tree Drawing Test; Psychometrics.

*Received 18 December 2012; accepted 20 February 2013*

### 1. Introduction

Projective techniques are indicated as reliable grounds for decisions made by a variety of institutions concerning a number of important areas in people’s lives. They were originally applied in clinical examinations and were one of the sources of information used when preparing psychological treatment programmes.

They were regarded then as a good means of establishing bonds and as a supplement to other clinical data: psychologists could verify the interpretations they based on them in the course of therapy. The situation is entirely different, for example, in court-ordered examinations and when preparing expert opinions for the needs of the courts. First of all, there is no opportunity to set the results of one examination against another and there is no room for rectifying an incorrect decision. Psychological interpretations of the collected data legitimise the decisions of others and it is impossible for psychologists to control this process. What is more, those who receive the decisions and opinions

\* This paper was written as a result of a panel discussion held during the “Conditions of Using Projection Methods in Psychological Diagnosis for Court Purposes” Conference, which was held on 16–17 November 2012 in Lublin.

treat them as final and irrevocable because they are based on reliable foundations. Rigorous demands are thus made of the reliability and validity of tools used to collect psychological data, while high levels of professional skill and knowledge are expected of the psychologists who apply them. And, though there is much imperfection in the use of diagnostic tools generally, it is especially possible in the case of projective techniques to point to many examples of errors in test administration, incorrect interpretation of results and clear cases of their over-interpretation. What is more, this concerns techniques of lesser repute more frequently than it does techniques such as the Rorschach test or TAT.

## 2. On the tree test and its evaluation

### 2.1. History and description

The creator of the Tree Test (German: Der Baumtest) was the psychologist Karol Koch (1906–1958). He was director of the Psycho-Technical Institute in Lucerne, Switzerland. The handbook he wrote was published in 1949 by Hans Huber Verlag under the title “Der Baumtest: Der Baumzeichenversuch als Psychodiagnostisches Hilfsmittel”. Elżbieta Renata Dajek’s Polish translation was first published by Erda in 1993 (the handbook for the test that is quoted here [20] is the second edition of this work). The test is distributed by PTP (Pracownia Testów Psychologicznych). The handbook specifies no required qualifications for applying the technique, while the distributor (PTP) requires a master’s degree in psychology.

The tree test was first presented by Karol Koch at an international conference of psychologists in Switzerland in 1947. The conference was followed by the publication of a collective work edited by Ludwig Klages and others [19], which contained Koch’s chapter. The tree test was first published as a single work in the same year [21].

The tool was intended by the author to assess personality based on the characteristics of a drawing of a tree. Children, young people and adults could all be assessed with the test. The task of the examinee begins and ends with the sentence, “please draw the most beautiful fruit tree you can. You can use the whole of the sheet of paper”. The examinee receives a sheet of white A4 paper, a medium-grade pencil and a rubber. If, in the opinion of the diagnostician the first drawing “does not correspond with the examinee” or is “not distinct enough” [20, p. 10], a second drawing may be requested. There are no instructions for diagnosticians

in the handbook regarding how they should behave while drawing is in progress other than to observe discreetly [20, p. 55]. In Koch’s view [20, p. 8] the test can be applied in all cases where “(...) for a variety of reasons it is necessary to do without graphology or projective tests, which are hard to learn and teach”.

It should be noted in passing that from the point of view of the projective hypothesis the tree test is not a projective test. This is because interpretation is based on the universal (the same for all) symbolism of the tree and its elements and not on discerning the projection of individual meanings onto the parts of the drawing (and this is not altered by the drawing on p. 13 [20]). No attempts are made in the procedure to individualise the interpretation by asking questions such as, “does the fact that the branches are bent mean something to you?”.

The drawing is assessed by reference to a large number of elements and the indicators that are based on them. These include the arrangement of the drawing on the surface of the paper, its size, the proportion of trunk to crown, the lifeline, the focus, the tilt of the tree, the roots and subsoil, the crown, the area of the crown and branches, the crown, the branches, additions to the drawing and other criteria. A very wide range of characteristics are attributed to the drawing’s author based on these elements, including maturity, level of intellectual development, disorder or delay in intellectual development, emotional inhibitions, empathy and aspirations.

The Wittgenstein index, named after the German neurologist Graf von Wittgenstein, merits a separate discussion. Koch does not describe this index in his book; John N. Buck [6] and Karen Bolander [5] are named as co-authors of this idea. This approach is very popular among practitioners in Poland, which is no doubt accounted for by the several widely-read publications that have advocated it [29, 33, 43]. Wittgenstein’s index is the ratio of the height of a tree in millimetres and the age of the patient in years and months, which is reported to make it possible to state their age at the time of an event (e.g. a traumatic experience) based on the height at which the symbol marking it, such as a hollow or a broken branch, is placed in the drawing. This may be regarded as a classic example of lending credibility to groundless speculation with the aid of numbers.

Neither Koch’s original handbook nor its Polish edition provide scales and neither describe the factors the test measures. Tables are provided in the handbook’s second section [20, pp. 56–96], in which interpretations of various forms of detail from drawings are given (along with a graphics template). There is no

information, however, on validation studies, no standards are provided and there is almost no academic literature on this tool.

In essence, all of the categories of interpretation give the impression of originating from popular rather than academic psychology. The author has freely selected the meanings of the concepts applied from a variety of sources, including non-academic ones, such as the poetry of Hiltbrunner or statements made by the painter Kandinsky. Among the psychologists Wittgenstein draws upon we find numerous advocates of graphology, as well as Rorschach, Lüscher and Jung. In the majority of cases there are no bibliographic references to specific publications and nor is there a list of literature cited (this is not a feature of the Polish version but of the German language original), which leaves the reader no choice but to take the author at his word. Koch draws on no research results whatever to substantiate the validity of these and other interpretations. He neglects to refer in any way to the need for the empirical verification of the psychological meanings he assigns to various features of drawings.

The tree test continues to be distributed in Poland in a form that has not changed since 1949, while the Polish language handbook is a straightforward translation of Koch's original work from 1949. As the translator, Elżbieta Renata Dajek, assures us in the introduction, the text is translated, "(...) with no short cuts or changes" [20, p. 5]. As a consequence, the first thing that strikes the reader is the shocking anachronism of the reasoning, and of the selection of arguments and interpretational categories, which are all too often judgemental (e.g. "aridity and aggression", "insolent" and "self-deceiving"). The author engages in stereotyping, for example by stating that deaf and hard of hearing people supposedly suffer from delayed intellectual development and uses the stigmatising term "retarded" [20, p. 23], which was long ago withdrawn from the lexicon of psychologists. We find further stereotypes in the description of young people on p. 27 or in the phrase, "peasant thinking" which here performs the function of a diagnostic category.

Another anachronism is that the example fragments from drawings in the second section of the handbook [20, pp. 56–96] originate from the period in which the test was being developed (1928–1952) and from a German-speaking area. It may be the case that the way contemporary Poles draw is analogous or similar – but this is a fairly risky thesis. This fact could be the cause of the test's invalidity. It would have to be extensively revised, and a Polish adaptation would need to be produced, for it to meet the requirements for validity. If Koch's book is to function as a contem-

porary test handbook rather than as an item belonging to psychology's past, the publisher of the Polish translation should (?) produce an adaptation of the tool for present-day Polish conditions.

The handbook includes fragments that are not so much unscientific as actively anti-scientific. They encourage the diagnosticians using the method to engage in practices that are the antithesis of scientific methodology and of any inferential rigour. Here are a few examples. The handbook permits the first set of results to be somehow invalidated and a further sample to be taken. And lo and behold on p.10 we read, "if someone has the impression that the first drawing does not correspond with the patient, or that the drawing is not sufficiently distinctive, the test may be repeated. Indeed it is possible to possess a whole series of drawings of trees and in many cases it is found that the second form of the drawing is truer". This is repeated in a modified form on p. 55 in the sentence, "if the drawing is excessively pedantic or unnatural, or if for other reasons the output is unsatisfactory, or if we wish to investigate other aspects and layers of personality, the test may be repeated many times. The instruction then reads, »please draw another fruit tree that is completely different from the one you have drawn«". The diagnostician therefore has the right according to criteria known only to him to select a certain fragment of the data and to deem it reliable while invalidating the remaining data and taking no account of it. This is a flagrant contradiction of scientific procedure. The second example involves encouraging the drawings to be interpreted with extreme subjectivity. The following paragraph appears on p. 36, "as with writing, a drawing may be received, interpreted as an intuitive whole, even without investigating the details one can obtain impressions of harmony, anxiety, emptiness, completeness, or a sense of hostility, prejudice (...). A person should passively consider the effect of a very large number of drawn trees, »contemplate them«, while at the same time examining them uncritically. In this way looking slowly turns to vision, the more important differences are recognised, the picture begins to differentiate and there follows a closer insight into the patient". This approach is exactly the type of false inference Meehl [25] warns of and that Lilienfeld et al. [22] termed impressionistic: in their view a method such as this cannot be investigated because it has no repeatable or systematic element. This lack of inferential rigour leads to serious over-interpretation.

## 2.2. Research

It can be seen from the handbook that the author has made a certain amount of effort to produce a type of normalisation for the test. On pp. 20–24 of the Polish translation of Koch's handbook there appear data concerning the age at which children draw a given detail in a particular way. The percentage of children in different age ranges who draw in this way is given further on. This is an important reference point for interpretation because the specific cognitive "unpunctuality" of a given feature of a drawing can be significant for the diagnosis of personality. In other words, what is typical at one age (assuming typical cognitive development) can at another be a potential indicator of "deeper levels of personality". Unfortunately, no information whatsoever is given regarding the nature and size of the sample or the method used to obtain the data. What is more, the discussion here is about children [20, pp. 20–23], while the test can also be used with young people and adults. This is not stated directly but can be inferred from various passages of the text [see 20, p. 20, 23, 25, 28, 32, 37]. There is no information whatsoever in the handbook on testing the reliability of the tool. These are therefore serious deficiencies concerning three indispensable psychometric attributes: validity, reliability and normalisation. The lack of any one of these three attributes would be sufficient to discredit the diagnostic technique. Koch himself writes [20, p. 38] of the technique's drawbacks and expresses the hope that it will undergo changes as a result of future experience and experiments. Yet the translation of the handbook into Polish with no modifications whatsoever in the version from 1952 does not bring these hopes of Koch's to fruition and instead renders the flaws of the technique glaringly visible.

These flaws had also been noticed earlier. The first review of the tree test appeared a few years after its publication [15] and was critical. The reviewer addressed the majority of the weaknesses set out above and stated in conclusion [15, p. 187] that "less fantasy and more facts" would be desirable. Koch's clinical interpretations have been contested in different periods [23], the original tree test procedure has been modified [40] and new indicators, such as a scale of emotional immaturity, have been constructed [39]. Various improvements [16] or even entirely new theoretical perspectives [27] have been proposed. One of the studies [3] confirmed the validity of the tree test in detecting intellectual problems; however the clinical value (indicators of neurosis) was not confirmed. A work containing passages on the tree test appeared at the beginning of this century in Italian [42]. It reviewed the pro-

jective techniques used in Italy. The authors concluded that, owing to the drawbacks concerning validity, reliability and normalisation for an Italian population, the techniques should only be applied ideographically but not for forensic or clinical diagnosis.

We may find information on personality tests in the book by Buros [7]. This survey of 513 personality tests and their reviews, which includes a fairly complete list of the publications from 1940–1968 devoted to the given test (in 1659 pages), includes only a short description of the tree test and two publications [7, p. 942]: Koch's book and an article by Alexander Tolor from 1957. Generally speaking the literature on the tree test is sparse. In the space of 65 years from 1947 to 2012 around 50–60 works have been published (according to a search in the basic psychology databases and to general searches of the EBSCO journals database, of the Internet and of other sources). That is not very many. A corresponding bibliography returns thousands of items for the Rorschach test and hundreds for TAT. The vast majority of the literature on the tree test is in languages other than English: only nine articles have been published in this language, of which three were found in the EBSCO databases. The publications in German appeared in the 1950s and 1960s and then ceased (a total of nine were found, including five in the EBSCO databases). The situation is similar with publications in French (there are seven, all of which were found in the EBSCO databases) and Spanish (two, both found in the EBSCO databases). These publications ceased in the 1970s, after which nothing is published on the tree test in these languages. Four items have appeared in Italian (all located in the EBSCO databases), of which one (referred to above) was published after 2000. One publication on the tree test in each of the following languages can be found in the EBSCO databases: Portuguese, Romanian, Czech, Hungarian, Russian and Chinese. There have also been a number of studies in Polish [2, 4, 31, 34], but these cannot be found in the EBSCO databases and are of uneven academic value.

In summarising the survey of the literature on the tree test it must be stated that it is modest both in terms of quality and quantity. The articles that have appeared are isolated: there is no correspondence between them, they are not cross referenced or they provide limited coverage. There are also no publications whatever in reputable international journals – not even one. On the other hand there is no lack of individual reports of obtaining truly very promising empirical results. However these are not followed up and no attempts are made by independent groups to replicate them.



It is also worth noting that much of the research mentioned here was conducted by doctors (psychiatrists, neurologists and also general practitioners), including in cooperation with specialists who were not psychologists, such as occupational therapists. This research is of limited value to psychology. It is striking that the tree test was (and still is) probably more popular among non-psychologists than among psychologists.

It may be stated in conclusion that the tree test does not meet contemporary criteria of reliability for diagnostic tools. There is an absence of scientific, theoretical grounds, empirical data and normalisation tests demonstrating the validity and reliability of the tool. There is no standard or even partially objectified system for interpreting the results. The diagnostician is encouraged to engage in practices that contradict the scientific method and that increase the risk of confirmation bias and other errors in diagnostic inference. The interpretative categories are anachronistic, value-laden and sometimes based on stereotypes. However, because of the absence of any coherent theoretical grounding, they do not include psychological knowledge. In light of the above it is necessary to state that in the present day the tree test is not a tool of psychological diagnosis and that if used as such it can be harmful.

It is perhaps worth quoting Karol Koch here [20, p. 36], “in the majority of tests interpretation is difficult and the shift from symptom to meaning, and from meaning to a description of personality or a diagnosis, depends very much on skill, experience and even on a sense of responsibility”.

### **3. Is the tree test actually often the basis for issuing decisions in various important areas of life?**

The authors of this article attempted to discover whether projective techniques are in fact often the basis for issuing decisions and, if so, what is the scale of this practice?

In the paper “Metodyka badań lekarskich i psychologicznych oraz kryteria i zasady oceny zdolności do pracy w ratownictwie górniczym” (Methodology of medical and psychological examinations and the criteria and principles for assessing the capacity to work in mine rescue) [26] the application of the tree test is recommended for the qualifying and periodic examinations administered to mine rescuers and mine rescue divers [26, p. 19]. As those who devised the methodology state, “the test makes it possible to ob-

tain information on the present mental state of those examined, on their predisposition to display or avoid certain behaviour, on their social relationships and on significant past events” [26, p. 21]. In addition to other tests, the Tree Test is therefore regarded as useful in the assessment of candidates for service in mine rescue. However no literature is indicated there on how to conduct examinations with the aid of the test, whereas a bibliographic note directing readers to test handbooks is provided for the other recommended tests. This may suggest that everyone has sufficient skills and knowledge to administer assessments using the tree test and to interpret the results.

Data concerning assessments of employees taken at random from 149 public psychological-pedagogical clinics from all provinces of Poland show that the tree test is used at 1.33% of these clinics [12]. This is a much lower proportion than for other types of examinations. In research involving 196 careers advisors it was found that [9] more than 15% of them used licensed psychological tools (those whose use requires an educational background in psychology). In reply to the question, “what licensed, standardised diagnostic tools, whose use requires an educational background in psychology, did you use in selecting candidates for a job, for training for employment or for an internship?” [9, pp. 128–129] more than 15% indicated the Tree Test (those asked were employees of district council employment offices). This is not a great number of people, yet the data still show the relative popularity of the Tree Test.

In a paper devoted to expert psychological opinions in cases of child sexual abuse, Sikorska-Koza [36] reported that the Tree Test was employed by more than 14% of the expert forensic psychologists who took part in the survey. Again, this is not an enormous proportion, but we may add that the tree test was the most popular of all the diagnostic tools (apart from the interview and observation) – all other diagnostic methods were used by less than 10% of the psychologists questioned.

Paweł Ostaszewski presents [28] the results of analyses of psychological opinions issued by Centres for Family Diagnosis and Consulting that were presented as evidence in 85 cases of applications for contact with children from people other than their parents concluded at a variety of district courts in 2000–2006. In addition to a psychological interview and observation, all of the opinions analysed also used projective techniques [28, p. 203], such as drawings and unspecified projective drawings, Rotter Incomplete Sentence Blanks, Koch’s Draw-A-Tree Test, Żelazowska’s Incomplete Sentence Trial, Sacks and Levy’s Sentence

Completion Test, Hess and Handel's Incomplete Sentence Series<sup>1</sup>, Kostrzewski's Incomplete Sentences Test, Thomas's Unfinished Story Test<sup>2</sup>, Corman's Draw-A-Family Test<sup>3</sup>, the Gille Film Test<sup>4</sup>, the Tree of Experiences Test<sup>5</sup> and the Lüscher Colour Test. Table I shows the percentage of cases the particular techniques were used.

TABLE I. FREQUENCY OF USE OF PROJECTIVES IN PSYCHOLOGICAL OPINIONS ( $N = 85$ ) FROM CENTRES FOR FAMILY DIAGNOSIS AND CONSULTING (BASED ON 28, P. 203)

Diagnostic techniques	Frequency of use [%]
Drawings and unspecified projective drawings	53.8%
Rotter Incomplete Sentence Blank	19.2%
Koch's Draw-a-Tree Test	14.1%
Żelazowska's Incomplete Sentence Trial	7.7%
Sacks and Levy's Sentence Completion Test	6.4%
Hess and Handel's Incomplete Sentence Series	2.6%
Kostrzewski's Incomplete Sentence Test	2.6%
Thomas's Unfinished Story Test	2.6%
Corman Draw-a-Family Test	4.7%
Gille Film Test	3.8%
Tree of Experiences Test	3.8%
Lüscher Colour Test	2.6%

Ostaszewski also writes [28, p. 205] that survey methods, which in his opinion are more objective, are used decidedly less often than projective methods. The use in such important matters of a method that lacks clear foundations and that is without empirical evidence in these areas is extremely unsettling. It is possible to accept the argument [28, p. 205] that it is difficult to apply survey methods to a population that is often made up of six-year-old children. However, as Ostaszewski writes [28, pp. 188–189], surveys are also used because their fundamental virtue is that “they have the capacity to eliminate deliberate attempts to mislead on the part of those surveyed (...) and that they present an all-round picture of the mental functioning of those surveyed”. Ostaszewski also describes [28,

p. 190] the results of analyses conducted by Wanda Stojanowska who, in 2000, investigated the psychological opinions issued in divorce cases by Centres for Family Diagnosis and Consulting. Psychological tests were used when writing opinions in 25% of the cases. The tests included methods such as Koch's Tree Test, M. Braun-Gałkowska's Draw-A-Family Test, Halpern and Machover's Draw-A-Person Test<sup>6</sup>, Sacks and Sidney's Sentence Completion Test<sup>7</sup>, Rotter's Free Association Test, Kostrzewski's Incomplete Sentences Questionnaire for Children, the Lüscher Colour Test, the Gille Film Test, Hess and Handel's Incomplete Sentence Series, Żelazowska's Incomplete Sentence Trial, Incomplete Sentences for Married Couples and the Two Houses Projective Test. In the report “Monitoring the work of selected Centres for Family Diagnosis and Consulting” presented at the conference organised in December 2009 at the headquarters of the Stefan Batory Foundation [13] it was shown that in an analysis of 81 opinions written in 2002–2009 as great a proportion as 64% employed various projective methods, whose status was vague and which were imprecisely or misleadingly named (they included Anna Frydrychowicz's Draw-A-Family Test, Koch's Tree Test, Żelazowska's Incomplete Sentence Trial for Children, Corman's Draw-A-Family Test, Sacks and Levy's Incomplete Sentence Trial for Adults, Kostrzewski's Incomplete Sentence Trial for Older Children and the Lüscher Colour Test [28, p. 37]). It should be noted that projective methods formed the only base for the opinion in 15% of opinions [28, p. 43]. The authors report that the tree test was used in approximately 28% of the opinions analysed [28, p. 42]. Even taking into account the non-random, deliberate selection of the opinions analysed (in 2008 alone Centres for Family Diagnosis and Consulting issued 23,045 opinions, including 9,247 in custody cases [28, p. 14]) it must provoke unease that projective techniques are used to such an extent in cases as controversial as these.

#### 4. On the need for reliability in tools and ways of assessing it

If something is reliable it is dependable and worthy of trust. In other words, psychological diagnostic tools are those on which the user can rely because someone

<sup>1</sup> From the forgotten work by Robert Hess and Gerald Handel [18]; the Polish translator is unknown.

<sup>2</sup> It is possible that this is the test of Madeleine Thomas [41]; the Polish translator is unknown.

<sup>3</sup> This probably comes from the work of Corman [10]; the Polish translator is unknown.

<sup>4</sup> It is possible that this is the test of René Gille [14]; quoted on p. 248 of Rembowski's book [33].

<sup>5</sup> The source could not be established.

<sup>6</sup> This is the Draw-A-Person Test developed by Karen Machover; see the following footnote.

<sup>7</sup> This of course refers to Sacks and Levy's test [35]. The low calibre of the authors of the opinions is already demonstrated by their not knowing the names of the authors of the tests whose results they draw upon.

has tested them earlier for reliability: are they sufficiently sensitive, specific and effective? [38, p. 50]. In addition to being reliable and valid, psychological tests should be standardised and have up-to-date norms that are consistent with the aims of the research (reference testing). They should also be objective (impartial in the sense of being independent of the diagnostician, non-biased) inter-subjectively testable and value free.

In the APA [1] standards describing the basic principles for applying tests no special category of “projective tests”, whose use would require separate regulation, is distinguished. They are treated in the same way as other psychological tests. It is only in the commentary to standard 12.12 that we find, “when using tests with unstructured answer formats, such as projective tests or informal assessments of behaviour, clear and objective scoring criteria that reduce the subjectivity of the assessment to a minimum should be used where possible” [1, p. 228]. There is therefore no reason to apply different criteria of reliability to projective tests than are applied to other diagnostic tools.

Several criteria for determining reliability have been adopted – mainly for use in forensic opinions. The first of these, which is known as the Daubert Standard, was developed in the United States [30, 37, 38]. Its original aim was to assure the accused that the examinations that were to form the basis of the opinion (evidence, expert opinion) would be performed according to clearly specified and tested criteria. Kirk Heilbrun [17] proposed other criteria that refer directly to the reliability of psychological tests. Douglas B. Marlowe [24], in creating a mixed scientific-legal model, developed and expanded them. In Poland, Alicja Czere-derecka [11] has devoted a great deal of attention to this matter. These proposals can be summarised in the form of the following list of criteria:

1. Is the theory (concept) behind the technique testable? Can it be empirically verified?
2. Does the theory (concept) behind the test refer to the area being analysed in the opinion, for which the test is the basis (directly or indirectly)? Has the area being analysed been subject to a test of the validity of the theory?
3. Is there empirical evidence, and if so what is it, that the given diagnostic technique is logically associated with the subject of the case or that it is widely used in such cases?
4. Has the diagnostic technique in question been described in a peer-reviewed expert publication? Has it been discussed at an academic conference?
5. Is the diagnostic technique in question accepted by scholars at the research centres that study the issue in question? Is it consistent with the present state

of knowledge in a given field or, more generally, is it consistent with the present state of science and learning?

6. Is the diagnostic technique in question available to be purchased along with a handbook and the appropriate norms? Are its basic psychometric parameters (including for the type of case and population concerned) known?
7. Have standards been established for the use of the diagnostic technique in question? Is there a text setting out the standard procedure for administering the test and for calculating and interpreting the results?
8. Is the diagnostic technique in question the correct one for the intended population and purpose?
9. Is the error rate obtained when this method is applied known? What is the technique’s measurement error? Is the possible influence of interference on the results of the diagnostic technique known? How susceptible is it to cheating?
10. Does the diagnostician have the appropriate knowledge and experience to administer the diagnostic test? Is this documented?
11. Has the diagnostician taken into account the probabilistic nature of the opinion expressed and is the opinion conclusive?

The question arises of which projective techniques meet these requirements and to what extent? It is certainly not the case that any one of them satisfies all of the requirements. But the situation is similar with regard to survey techniques. We therefore return to the question: what makes a diagnostic tool reliable? It would appear that continual validity testing is the decisive factor. There is a widespread but mistaken conviction that the correctness of the methodological procedures is only important in the preparation phase of the test and when it is being readied for market and, it follows, that once it has been published the psychological test is a finished product (tool). Yet what happens once it has been published is equally important: Will the norms be continually updated? Will there be academic research into the areas of its validity? Diagnostic tools that are written about in the scientific literature are a good deal more credible than tools that are the subject of either a handful of publications or none at all (apart from the test handbook). In 1940–2001 close to 14,000 scholarly articles appeared [8] on the subject of MMPI and MMPI-2, which gives an annual average of 226 ( $\pm$  116) publications. In the 1950s an annual average of 111.6 articles were published on the Rorschach Test and of 26.1 on MMPI, while in 1970–1979 an annual average of 64.6 articles were published on the Rorschach Test and of 115.4 articles on MMPI [32].

This demonstrates the scale of the problem. It is worth adding – based on an analysis of the literature cited in Buros's review of tests [7] – that it can be clearly seen that beginning from 1960 there has been a systematic decline in the interest of researchers in projective techniques and a concurrent rise in their interest in survey techniques<sup>8</sup>.

The value of a psychological test is thus decided by observing psychological standards – during both development and practical use – and by the number of academic articles written about it. The question of basing diagnosis on scientific evidence as well as of the specific monitoring of the commercial market for diagnostic tools is important: on the one hand this is because the emphasis on analysing the reliability of psychological diagnostic tools is growing steadily stronger and on the other because potential clients' lack of knowledge favours a certain 'tabloidisation' of tests.

## References

1. APA, Standardy dla testów stosowanych w psychologii i pedagogice, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2007.
2. Basińska B., Czuchnowska H., Test Drzewa w diagnozie adaptacji do choroby nowotworowej, *Psychoonkologia* 2000, 7, 39–48.
3. Bernet A., Research regarding the Draw-A-Tree Test, *Revue de Psychologie Appliquée* 1971, 21, 101–122.
4. Bieniewicz-Wolak M., Wyras przeżytej traumy w rysunku drzewa a nasilenie przeżyć dysocjacyjnych, *Rocznik Naukowy Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy* 2010, 5, 19–29.
5. Bolander K., Assessing personalities through tree drawings, Basic Books, New York 1977.
6. Buck J. N., The H-T-P technique: A qualitative and quantitative scoring manual, *Journal of Clinical Psychology* 1948, 4, 317–396.
7. Buros O. K. [ed.], Personality. Tests and reviews, The Gryphon Press, New Jersey 1970.
8. Butcher J. N., Atlis M., Hahn J., The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) [in]: Comprehensive handbook of psychological assessment, vol. 2: Personality assessment, Hilsenroth M. J., Segal D. L., Hersen M. [eds.], John Wiley & Sons, New York 2003.
9. Chrościcka A., Płachecki T., Dymek K. [et al.], Kompleksowa analiza pomocy pracodawcom w doborze pracowników, Agencja Badań Rynku OPINIA, Warszawa 2008.
10. Corman L., The Family Drawing Test in the study of pathological reactions caused by sibling rivalry, *Journal de Médecine de Nantes* 1964, 21, 229–247.
11. Czerederecka A., Test Rorschacha w psychologicznej ekspertyzie sądowej: wykorzystanie systemu Exnera, Wydawnictwo IES, Kraków 2006.
12. Dzielnicka E., Wierzchowska-Szymanek A., Poradnictwo zawodowe w poradniach psychologiczno-pedagogicznych. Raport z badań, Ecorys Polska Sp. z o.o., Warszawa 2008.
13. Dzierżęga I., Borkowski M., Gujska B. [et al.], Monitoring pracy wybranych Rodzinnych Ośrodków Diagnostyczno-Konsultacyjnych, Stowarzyszenie Centrum Praw Ojca i Dziecka, Warszawa 2009.
14. Gille R., Le test-film; un instrument pour la mesure objective du niveau de maturité affective et de certains traits du comportement, Presses universitaires de France, Paris 1959.
15. Gleser G. C., Review of "The Tree Test", *Journal of Educational Psychology* 1954, 45, 186–187.
16. Harsányi I., Donáth B. G., The tree-drawing-test, *Magyar Pszichológiai Szemle* 1978, 35, 3–18.
17. Heilbrun K., The role of psychological testing in forensic assessment, *Law and Human Behavior* 1992, 16, 257–272.
18. Hess R., Handel G., Family Worlds: A psychosocial approach to family life, University of Chicago Press, Chicago 1959.
19. Klages L. [eds., et al.], Le diagnostic du caractère, Presses Universitaires de France, Paris 1949.
20. Koch Ch., Test Drzewa. Test rysunku drzewa jako metoda psychodiagnostyczna, Erda, Warszawa 1997.
21. Koch K., Der Baumtest, Verlag Hans Huber, Bern 1949.
22. Lillienfeld S. O., Wood J. M., Garb H. N., Status naukowy technik projekcyjnych, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2002.
23. Loreto D., Koch's tree test: Some theoretical and applied situations, *Neuropsychiatria* 1973, 29, 59–64.
24. Marlowe D. B., A hybrid decision framework for evaluating psychometric evidence, *Behavioral Sciences & the Law* 1995, 13, 207–228.
25. Meehl P. E., The cognitive activity of the clinician, *American Psychologist* 1960, 15, 19–27.
26. Metodyka badań lekarskich i psychologicznych oraz kryteria i zasady oceny zdolności do pracy w ratownictwie górniczym, Centralna Stacja Ratownictwa Górniczego w Bytomiu, Bytom 2008.
27. Noss I. N., Diagnostic capacities of the graphic test "The tree" by K. Koch, *Voprosy Psychologii* 2006, 2, 148–156.

<sup>8</sup> According to figures from the "Publish or Perish" programme, in 1940–2010 the combined number of citations of articles on the Rorschach Test was 104,632, while there were 195,558 citations for MMPI/MMPI2.



28. Ostaszewski P., Opinia rodzinnego ośrodka diagnostyczno-konsultacyjnego w sprawach o ustalenie kontaktów dziecka z innymi osobami niż rodzice, [w:] Prawo w działaniu, t. 4: Sprawy cywilne, Holewińska-Lapińska E. [ed.], Instytut Wymiaru Sprawiedliwości, Oficyna Naukowa, Warszawa 2008.
29. Oster G. D., Gould, P., Rysunek w psychoterapii, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2001.
30. Paluchowski W. J., Diagnoza oparta na dowodach empirycznych – czy potrzebny jest „polski Buros”?, *Roczniki Psychologiczne* 2010, 13, 7–28.
31. Pędziwiatr H., Czajkowska M., The results of Koch's Tree Test investigation for persons suffering from illnesses connected with intervals of the cervical spine, *Polish Journal of Applied Psychology* 2006, 4, 59–72.
32. Polyson J., Peterson R., Marshall Ch., MMPI and Rorschach: Three decades of research, *Professional Psychology: Research and Practice* 1986, 17, 476–478.
33. Rembowski J., Metoda projekcyjna w psychologii dzieci i młodzieży, Państwowe Wydawnictwo Naukowe, Warszawa 1986.
34. Rutkowski K., Turkot A., Kurek-Rusin A., Reakcja przewlekłą psychozą na uraz psychiczny doznany w dzieciństwie – opis przypadku, *Psychoterapia* 2007, 140, 75–84.
35. Sacks J. M., Levy S., The Sentence Completion Test, [in:] *Projective psychology: Clinical approaches to the total personality*, Abt L. E., Bellak L. [eds.], Alfred A. Knopf, New York 1950.
36. Sikorska-Koza A., Ocena psychologicznych aspektów wiarygodności zeznań małoletnich świadków w praktyce biegłych sądowych psychologów na podstawie analizy spraw karnych dotyczących wykorzystania seksualnego dzieci, *Dziecko Krzywdzone* 2010, 1, 70–89.
37. Stemplewska-Żakowicz K., O rzeczach widywanych na obrazkach i opowiadanych o nich historiach. TAT jako metoda badawcza i diagnostyczna, *Academica*, Warszawa 2004.
38. Stemplewska-Żakowicz K., Paluchowski J. W., Podstawy diagnozy psychologicznej, [in:] *Psychologia. Podręcznik akademicki*, vol. 2, Strelau J., Doliński D. [eds.], Wydawnictwo Psychologiczne, Gdańsk 2008.
39. Stora A., La personnalité a travers le test de l'arbre, *Bulletin de Psychologie* 1964, 17, 1–181.
40. Stora R., L'arbre de Koch; test original et test modifié, *Enfance* 1948, 1, 327–344.
41. Thomas M., Méthode des histoires à compléter, pour le dépistage des complexes et des conflits affectifs infantiles, *Archives de Psychologie* 1937, 26, 209–284.
42. Tressoldi P. E., Barilani C., Pedrabissi L., Lo stato (preoccupante) delle tecniche proiettive per l'età evolutiva in Italia, *Psicologia Clinica dello Sviluppo* 2004, 8, 9–28.
43. Wach E., Możliwości wykorzystania Testu Drzewa w ekspertyzie psychologicznej, [in:] *Techniki projekcyjne w psychologicznej ekspertyzie sądowej*, Czerederecka A., Jaśkiewicz-Obydzińska T. [eds.], Wydawnictwo IES, Kraków 1996.

---

**Corresponding author**

Prof. dr hab. Władysław Jacek Paluchowski  
Uniwersytet im. Adama Mickiewicza  
Instytut Psychologii  
ul. A. Szmarzewskiego 89  
PL 60-568 Poznań  
e-mail: wladyslaw.paluchowski@amu.edu.pl

---

## WIARYGODNOŚĆ TECHNIK PROJEKCYJNYCH JAKO NARZĘDZI DIAGNOZY PSYCHOLOGICZNEJ. CZĘŚĆ 2: TEST DRZEWA\*

### 1. Wprowadzenie

Techniki projekcyjne wskazywane są jako wiarygodna podstawa do podejmowania przez różne instytucje decyzji dotyczących niektórych ważnych sfer życia człowieka. Stosowane pierwotnie podczas badań klinicznych, były one częścią informacji do przygotowania planu interwencji psychologicznej. Traktowano je wówczas jako dobry sposób budowania więzi, jako uzupełnienie innych danych klinicznych – psycholog mógł swoje oparte na nich interpretacje weryfikować w trakcie terapii. Jednak zupełnie inna sytuacja występuje na przykład w badaniach orzecznich i przy opracowywaniu opinii dla potrzeb sądowych. Po pierwsze, nie ma okazji do monitorowania efektów takiego a nie innego orzeczenia oraz nie ma miejsca na korektę nietrafnej opinii. Psychologiczne interpretacje zebranych danych legitymizują cudze decyzje i kontrola tego procesu przez psychologa jest niemożliwa. Ponadto odbiorcy opinii i orzeczeń traktują je jako ostateczne i nieodwołalne, bo oparte na wiarygodnych podstawach. Wszystko to nakłada na narzędzia zbierania danych psychologicznych wysokie wymagania co do ich rzetelności i trafności, a od stosujących te narzędzia psychologów oczekuje się wysokich kompetencji zawodowych. I choć obserwuje się wiele niedoskonałości w posługiwaniu się instrumentarium diagnostycznym, to szczególnie podczas używania technik projekcyjnych wskazać można wiele przykładów niepoprawnej administracji testów, nietrafnej interpretacji wyników badania oraz oczywistych ich nadinterpretacji. Co więcej – rzadziej dotyczy to takich technik, jak test Rorschacha czy TAT, a częściej technik o mniejszym ciężarze gatunkowym.

### 2. O Teście Drzewa i jego ocenie

#### 2.1. Historia i opis testu

Twórcą Testu Drzewa (niem. Der Baumtest) jest psycholog Karol Koch (1906–1958). Był on kierownikiem Instytutu Psychotechnicznego w Lucernie w Szwajcarii. Podręcznik jego autorstwa ukazał się w 1949 roku w wydawnictwie Hans Huber Verlag pod tytułem „Der Baum-

test: Der Baumzeichenversuch als Psychodiagnostisches Hilfsmittel”. Polskie tłumaczenie Elżbiety Renaty Dajek ukazało się w wydawnictwie Erda po raz pierwszy w 1993 roku (cytowany dalej podręcznik do testu [20] to drugie wydanie tejże pracy). Dystrybutorem testu jest Pracownia Testów Psychologicznych PTP. Podręcznik nie określa żadnych wymaganych kwalifikacji do stosowania tej techniki, dystrybutor (Pracownia Testów Psychologicznych) wymaga dyplomu ukończenia studiów psychologicznych.

Test Drzewa po raz pierwszy był prezentowany przez Karola Kocha na międzynarodowym zjeździe psychologów w Szwajcarii w roku 1947. Po konferencji została opublikowana praca zbiorowa pod redakcją Ludwiga Klagesa i innych [19], w której zamieszczono rozdział Kocha. Samodzielnej publikacji Test Drzewa doczekał się w tym samym roku [21].

Narzędzie w zamyśle autora jest przeznaczone do diagnozy osobowości na podstawie charakterystyki rysunku drzewa. Osobami badanymi za pomocą tego testu mogą być dzieci, młodzież i dorośli. Zadanie, przed jaki staje osoba badana zamyka się w zdaniu: „proszę, narysuj drzewo owocowe, tak ładnie, jak potrafisz. Możesz użyć całej tej kartki”. Osoba diagnozowana otrzymuje białą kartkę papieru w formacie A4 oraz średnio miękkie ołówki i gumkę. Jeśli zdaniem diagnosty pierwszy rysunek „nie koresponduje z osobą badaną” albo „różnicuje zbyt mało” [20, s. 10], to można poprosić o powtórne narysowanie drzewa. W podręczniku nie ma wskazówek, jak diagnosta powinien zachować się podczas rysowania, jest tylko sugestia, że dobrze, aby dyskretnie obserwował [20, s. 55]. Zdaniem Kocha [20, s. 8] ma on zastosowanie wszędzie tam, „(...) gdzie z różnych przyczyn muszą obejść się bez używania grafologii i testów projekcyjnych, których trudno się nauczyć”.

*Nota bene* z punktu widzenia hipotezy projekcyjnej, Test Drzewa (dalej oznaczany skrótem TD) nie jest testem projekcyjnym, ponieważ interpretacja opiera się na uniwersalnej (takiej samej dla wszystkich) symbolice drzewa i jego elementów, nie zaś na dopatrywaniu się rzutowania indywidualnych znaczeń na elementy rysunku (i nie zmienia tego diagram ze s. 13 [20]). W całej procedurze badania nie ma żadnych prób indywidualizacji interpretacji w postaci na przykład pytań „czy to coś dla ciebie znaczy, że te gałęzie są pochylone?” lub podobnych.

W ocenie rysunku bierze się pod uwagę wielką liczbę różnych jego elementów i opartych na nich wskaźników, w tym rozmieszczenie rysunku na płaszczyźnie kartki, wielkość rysunku, proporcje narysowanego pnia

\* Niniejsza praca powstała w wyniku dyskusji panelowej prowadzonej podczas konferencji „Warunki wykorzystywania metod projekcyjnych w psychologicznej diagnozie dla potrzeb sądu”, która odbyła się w dniach 16–17 listopada 2012 roku w Lublinie.

i korony, tzw. linia życia, punkt ciężkości, wychylenie drzewa, korzenie i podłoże, pień, powierzchnia pnia i gałęzi, korona, gałęzie, uzupełnienia rysunku oraz inne kryteria. Z tych elementów rysunku wyciągane są wnioski na temat bardzo różnorodnych właściwości jego autora, w tym dojrzałości osobowości, poziomu rozwoju intelektualnego i opóźnień w nim, zahamowania emocjonalnego, empatii, dążeń i in.

Na odrębne omówienie zasługuje tzw. wskaźnik Wittgensteina, nazwany tak od nazwiska niemieckiego neurologa Graf von Wittgensteina. Koch w swojej książce nie opisuje tego wskaźnika, jako współautorów tego pomysłu wskazuje się Johna N. Bucka [6] i Karen Bolander [5]. Jest to podejście bardzo popularne wśród praktyków w Polsce, do czego zapewne przyczyniło się kilka poczytnych publikacji je promujących [29, 33, 43]. Wskaźnik Wittgensteina to iloraz wysokości drzewa w milimetrach przez wiek życia w latach i miesiącach, podobno pozwalający określić wiek osoby badanej w chwili wystąpienia danego wydarzenia (np. urazowego przeżycia) na podstawie wysokości umieszczenia symbolu określającego dane wydarzenie (np. dziupli, złamanego konaru) na rysunku. Można to uznać za klasyczny przykład uwiarygodniania bezpodstawnej spekulacji za pomocą liczb.

W oryginalnym podręczniku Kocha i jego polskiej edycji ani nie zostały podane skale, ani też nie opisano konstruktów (czynników), które test mierzy. W drugiej części podręcznika [20, s. 56–96] znajdują się tzw. tablice, w których podano interpretacje różnych form detali rysunku (wraz z graficznym wzorcem). Brak natomiast informacji o badaniach walidacyjnych, brak norm, praktycznie brak też literatury akademickiej na temat tego narzędzia.

Wszystkie w zasadzie kategorie interpretacji sprawiają wrażenie pochodzących z psychologii potocznej, a nie naukowej. Znaczenia stosowanych pojęć autor czerpał swobodnie z różnych, także pozanaukowych źródeł, np. z poezji Hiltbrunnera czy też wypowiedzi malarza Kandinsky'ego. Spośród psychologów przywoływani są liczni reprezentanci grafologii oraz Rorschach, Lüscher i Jung. W większości przypadków nie ma odsyłaczy bibliograficznych do konkretnych publikacji ani też spisu literatury cytowanej (nie jest to cecha wersji polskiej, lecz niemieckojęzycznego oryginału), co sprawia, że czytelnik musi wierzyć autorowi na słowo. Koch nie uzasadnia trafności tych i pozostałych interpretacji poprzez powołanie się na wyniki jakichkolwiek badań. Nie odnosi się w żaden sposób do potrzeby empirycznego zweryfikowania psychologicznych znaczeń, które przypisuje różnym cechom rysunku.

Test Drzewa jest obecnie w Polsce rozprowadzany w postaci, która nie uległa zmianie od roku 1949. Podręcznik w języku polskim stanowi proste tłumaczenie oryginalnej pracy Kocha z 1949 roku. Jak we wprowadzeniu zapewnia tłumaczka, Elżbieta Renata Dajek, tekst

przełożono „bez jakichkolwiek skrótów i zmian” ([20, s. 5]. W konsekwencji pierwszą rzeczą, która rzuca się w oczy czytelnikowi, jest szokująca anachroniczność wywodów oraz doboru argumentów i kategorii interpretacyjnych, które nader często mają charakter oceniający (np. „jałowość i agresja”, „bezczelny”, „zakłamaný w stosunku do siebie”). Autor posługuje się stereotypami, np. twierdzi, jakoby osoby niesłyszące i niemówiące były opóźnione w rozwoju umysłowym i stosuje dawno wycofany ze słownika psychologów stygmatyzujący termin „niedorozwinięci” [20, s. 23]. Kolejne stereotypy znajdujemy np. w opisie młodych ludzi na s. 27 czy we frazie „chłopskie myślenie”, która pełni tu funkcję kategorii diagnostycznej [20, s. 57].

Innym anachronizmem jest, że wzorcowe fragmenty rysunków znajdujące się w drugiej części podręcznika [20, s. 56–96] pochodzą z czasów tworzenia tego testu, czyli lat 1928–1952, z obszaru niemieckojęzycznego. Być może sposób rysowania współczesnych Polaków jest analogiczny lub zbliżony, ale jest to teza dość ryzykowna. Fakt ten może być potencjalną przyczyną nie trafności testu. Aby sprostać wymaganiu trafności, Test Drzewa musiałby zostać poddany głębokiej rewizji oraz przejść polską adaptację. Jeśli książka Kocha nie ma być pozycją z historii psychologii, ale współcześnie używanym podręcznikiem testowym, to wydawca polskiego tłumaczenia powinien (?) dokonać dostosowania narzędzia do aktualnych polskich warunków.

Treści zawarte w podręczniku zawierają fragmenty o charakterze nie tyle nawet nienaukowym, co aktywnie antynaukowym. Zachęcają diagnostów posługujących się tą techniką do praktyk będących przeciwieństwem naukowej metodologii i jakiegokolwiek dyscypliny wnioskowania. Poniżej kilka przykładów. Podręcznik zezwala na to, żeby pierwsze wyniki niejako unieważnić i zebrać kolejną próbkę. Oto na s. 10 czytamy: „Jeśli ktoś ma wrażenie, że pierwszy rysunek nie koresponduje z osobą badaną, albo że rysunek różnicuje zbyt mało, to test można powtórzyć. Rzeczywiście jest możliwe posiadanie całej serii rysunków drzew i w wielu przypadkach stwierdzono, że ta druga forma rysunku jest prawdziwsza”. Jest to powtórzone w zmodyfikowanej formie na s. 55 w zdaniu: „Przy rysowaniu zanadto pedantycznym lub nienaturalnym albo z innych względów niedostatecznej produkcji albo jeśli chcemy zbadać inne aspekty i warstwy osobowości, test można powtarzać wielokrotnie. Instrukcja brzmi wtedy: »proszę, narysuj inne drzewo owocowe, zupełnie różne od tego, które narysowałeś«”. Diagnosta ma więc prawo wedle siebie tylko znanych kryteriów wybrać pewien fragment danych i uznać go za miarodajny, a pozostałe dane unieważnić, nie brać ich pod uwagę. Jest to jaskrawe zaprzeczenie postępowania naukowego. Drugi przykład to zachęcanie do skrajnie subiektywnego interpretowania rysunku. Na s. 36 widnieje następujący ustęp: „Podobnie jak pismo, rysunek może

być odbierany jako całość intuicyjnie, nawet bez badania szczegółów można odbierać wrażenie harmonii, niepokoju, pustki, nagości albo pełni, albo wrażenie wrogości, uprzedzenia (...). Człowiek powinien pasywnie rozważać skutek bardzo wielkiej liczby narysowanych drzew, »kontemplować je«, jednocześnie przyglądać się im bez żadnego krytycznego nastawienia. Tak pomału patrzenie przechodzi w widzenie, rozpoznaje się ważniejsze różnice, obraz zaczyna się różnicować, następuje bliższe poznanie badanego». Takie podejście jest właśnie tym rodzajem błędnego wnioskowania, przed jakim ostrzegał Paul Meehl [25], a który Scott O. Lilienfeld i współpracownicy [22] określili mianem impresyjnego i uznali, że nie można poddać go badaniom, ponieważ nie ma w nim żadnego powtarzalnego, systematycznego elementu. Ten brak dyscypliny wnioskowania prowadzi do poważnych nadinterpretacji.

## 2.2. Badania

W podręczniku widać pewien trud podjęty przez autora w celu stworzenia pewnego rodzaju normalizacji testu. Na s. 20–24 podręcznika Kocha w polskim tłumaczeniu umieszczono dane co do tego, w jakim wieku dzieci rysują dany detal w ten lub inny sposób. Dalej podano też, jaki odsetek dzieci w różnych przedziałach wiekowych rysuje w ten sposób. To ważny punkt odniesienia do interpretacji, gdyż istotna dla osobowościowej diagnozy może być swoista „niepunktualność” poznawcza danej cechy rysunku. Innymi słowy to, co typowe w jednym wieku, w innym może potencjalnie być (przy założeniu typowego rozwoju poznawczego) wskaźnikiem „głębszych warstw osobowości”. Niestety nie ma żadnych informacji o tym, od jakiej próby, jak dużej i w jaki sposób uzyskano te dane. Ponadto mowa tam [20, s. 20–23] o dzieciach, a test ten stosowany może być także u młodzieży i osób dorosłych (nie podano tego wprost, ale wynika to z różnych fragmentów tekstu [por. 20, s. 20, 23, 25, 28, 32, 37]). Na temat badania rzetelności narzędzia w podręczniku nie ma żadnych informacji. Są to poważne braki dotyczące trzech kluczowych jakości psychometrycznych – trafności, rzetelności i normalizacji. Każdy z tych braków sam w sobie może zdecydować o zdyskredytowaniu techniki diagnostycznej. Sam Koch mówi [20, s. 38] o mankamentach techniki i wyraża nadzieję, że przejdzie ona zmiany pod wpływem przyszłych doświadczeń. Jednak przetłumaczenie na język polski bez żadnych modyfikacji podręcznika w wersji z 1952 roku nie realizuje tych nadziei Kocha i czyni wady techniki jaskrawo widocznymi.

Dostrzegano te wady także i wcześniej. Pierwsza recenzja TD ukazała się kilka lat po jego opublikowaniu [15] i była krytyczna. Recenzentka wypunktowała większość z poruszonych wyżej słabości pracy i w konkluzji [15, s. 187] stwierdziła, że można by sobie życzyć „mniej

fantazji, więcej faktów”. W różnych okresach polemizowano z Kochem w kwestiach dotyczących klinicznej interpretacji [23], modyfikowano oryginalną procedurę TD [40], konstruowano nowe wskaźniki, np. skalę niedojrzałości emocjonalnej [39]. Proponowano też różne udoskonalenia [16] lub nawet całkowicie nowe ujęcie teoretyczne [27]. W jednym z badań [3] potwierdzono trafność TD w wykrywaniu problemów intelektualnych, natomiast wartość kliniczna (wskaźniki nerwic) nie została potwierdzona. Już w nowym stuleciu ukazała się praca w języku włoskim [42], w której znalazł się fragment na temat TD. W pracy dokonano przeglądu technik projekcyjnych stosowanych we Włoszech. W konkluzji autorzy stwierdzają, że z powodu mankamentów dotyczących trafności, rzetelności i norm dla populacji włoskich, techniki te powinny być stosowane jedynie idiograficznie, natomiast nie do celów związanych z diagnozą sądową czy kliniczną.

W książce Oskara Krisena Burosa [7] można szukać informacji na temat testów osobowości. Ten przegląd 513 testów osobowości i ich recenzji oraz w miarę kompletny spis publikacji poświęconych danemu testowi opublikowanych w latach 1940–1968 (wszystko na 1659 stronach) zawiera jedynie krótki opis TD i dwie publikacje [7, s. 942]: książkę Karola Kocha i artykuł Alexandra Tolora z 1957 roku. Generalnie piśmiennictwo na temat Testu Drzewa jest raczej ubogie. Na przestrzeni 65 lat (od roku 1947 do 2012) ukazało się około 50–60 prac (według wyszukiwania w podstawowych bazach psychologicznych i ogólnych w bazie czasopism EBSCO, w zasobach Internetu oraz innych źródeł). To niewiele. Analogiczna bibliografia dla Testu Rorschacha liczy tysiące, a dla TAT setki pozycji. Ogromna większość literatury na temat TD jest w językach innych niż angielski, w którym wydano tylko 9 prac (z czego 3 są notowane w bazach EBSCO). Publikacje niemieckojęzyczne ukazywały się w latach 50. i 60. XX wieku, po czym zanikły (w sumie znaleziono ich 9, w tym 5 w bazach EBSCO). Podobnie z publikacjami w języku francuskim (jest ich 7, wszystkie w EBSCO) i hiszpańskim (2 – obie w EBSCO). Publikacje te kończą się na latach 70. XX wieku, potem już nie publikowano na temat TD w tych językach. Po włosku opublikowano 4 prace (wszystkie w EBSCO), z czego jedną (wyżej wspomnianą) po roku 2000. W EBSCO można też znaleźć po 1 pracy opublikowanej w języku portugalskim, rumuńskim, czeskim, węgierskim, rosyjskim i chińskim. Po polsku także powstało kilka prac [2, 4, 31, 34], jednak nienotowanych w EBSCO i o różnej wartości naukowej.

Podsumowując przegląd piśmiennictwa na temat TD, trzeba stwierdzić, że jest ono skromne zarówno ilościowo, jak i jakościowo. Pozycje, które się ukazały, są odosobnione, nie ma między nimi korespondencji ani wzajemnych odwołań lub też istnieją one w niewielkim zakresie. Nie ma też żadnych – dosłownie ani jednej –



publikacji w renomowanych pismach o zasięgu międzynarodowym. Z drugiej strony nie brakuje pojedynczych doniesień o uzyskaniu doprawdy bardzo obiecujących wyników empirycznych – nie ma jednak kontynuacji tych badań ani prób replikacji przez niezależne zespoły.

Warto też zauważyć, że wiele badań tu wspomnianych było prowadzonych przez lekarzy (psychiatrów, neurologów, ale też lekarzy ogólnych), ewentualnie we współpracy z innymi specjalistami niepsychologami (np. terapeutami zajęciowymi). Wartość tych badań dla psychologii jest ograniczona. Zastanawiające, że TD był i wciąż jest bodaj bardziej popularny wśród niepsychologów niż wśród psychologów.

W konkluzji można stwierdzić, że TD nie spełnia współczesnych kryteriów wiarygodnego narzędzia diagnostycznego. Brak jest naukowych podstaw teoretycznych, jak również danych empirycznych dowodzących trafności i rzetelności narzędzia, brak też badań normalizacyjnych. Nie ma standardowego i choćby częściowo zobiektywizowanego systemu interpretacji wyników. Diagnosta jest zachęcany do praktyk sprzecznych z naukową metodologią, które zwiększają ryzyko wystąpienia myślenia konfirmacyjnego i innych błędów we wnioskowaniu diagnostycznym. Kategorie interpretacyjne są anachroniczne, oceniające, niekiedy oparte na stereotypach, natomiast – z racji braku spójnych podstaw teoretycznych – nie obejmują wiedzy psychologicznej. W świetle powyższego trzeba stwierdzić, że współcześnie TD nie jest psychologicznym narzędziem diagnostycznym, a stosowany w tej funkcji może być szkodliwy.

Warto może w tym miejscu zacytować Karola Kocha [20, s. 36]: „W większości testów interpretacja jest trudna i krok od symptomu do znaczenia i od znaczenia do opisu osobowości albo diagnozy jest bardzo zależny od zdolności, doświadczenia, a nawet od poczucia odpowiedzialności”.

### 3. Czy faktycznie Test Drzewa jest często podstawą wydawania decyzji w różnych ważnych sferach życia?

Autorzy tego artykułu szukali informacji, czy faktycznie techniki projekcyjne są często podstawą do wydania decyzji oraz jaka jest skala tego problemu?

W pracy „Metodyka badań lekarskich i psychologicznych oraz kryteria i zasady oceny zdolności do pracy w ratownictwie górniczym” [26] do badań kwalifikacyjnych i okresowych ratowników górniczych, ratowników górniczych nurków [26, s. 19] poleca się stosowanie TD. Jak piszą autorzy metodyki, „Test umożliwia uzyskanie informacji o obecnym stanie psychicznym osoby badanej, jak i o jej predyspozycjach do przejawiania lub unikania pewnych zachowań, oraz o jej relacjach społecz-

nych i istotnych zdarzeniach z przeszłości” [26, s. 21]. Oprócz innych testów TD jest więc uznany za przydatny do oceny kandydatów do służby w ratownictwie górniczym. Nie ma tam jednak wskazanej żadnej literatury dotyczącej prowadzenia badania za pomocą tego testu, choć dla innych zalecanych testów podano notkę bibliograficzną podręczników testowych. To może sugerować, że każdy ma wystarczające kompetencje do badania TD i interpretowania jego wyników.

Dane z badania pracowników wylosowanych 149 publicznych poradni psychologiczno-pedagogicznych ze wszystkich województw Polski pokazują, że TD wykorzystywano w 1,33% poradni [12]. To zdecydowanie mniej niż przy innych badaniach. W badaniach 196 doradców zawodowych okazało się [9], że z licencjonowanych narzędzi psychologicznych (takich, których stosowanie wymaga wykształcenia psychologicznego) korzystało ponad 15% osób. W odpowiedzi na pytanie „jakich licencjonowanych, wystandaryzowanych narzędzi diagnostycznych, których stosowanie wymaga wykształcenia psychologicznego, używał(a) P., dokonując doboru kandydatów na stanowisko pracy, przygotowania zawodowego lub stażu?” [9, s. 128–129], ponad 15% wskazywało na TD (byli to pracownicy powiatowych urzędów pracy). Nie jest to wielka liczba osób, ale jednak dane te pokazują relatywną popularność TD.

W pracy poświęconej opiniom psychologicznym w sprawach o seksualne wykorzystanie dziecka Sikorska-Koza [36] podaje, że TD był użyty przez ponad 14% uczestniczących w sondażu biegłych sądowych psychologów. Znowu nie jest to bardzo dużo, jednak dodajmy, że TD był najbardziej popularny spośród wszystkich narzędzi diagnostycznych (oprócz wywiadu i obserwacji) – każda inna technika diagnostyczna była stosowana przez mniej niż 10% pytanych psychologów.

Paweł Ostaszewski przedstawia [28] wyniki analizy psychologicznych opinii Rodzinnych Ośrodków Diagnostyczno-Konsultacyjnych (RODK), które były dowodami w 85 sprawach o ustalenie kontaktów dzieci z osobami innymi niż ich rodzice, a zakończonych w latach 2000–2006 i pochodzących z różnych sądów rejonowych. Wśród wszystkich analizowanych opinii, oprócz rozmowy psychologicznej i obserwacji, wykorzystywano też techniki projekcyjne [28, s. 203]. Wśród wszystkich analizowanych opinii, oprócz rozmowy psychologicznej i obserwacji, wykorzystywano [28, s. 203] takie techniki projekcyjne, jak rysunki i nieopisane dokładniej rysunki projekcyjne, Test Niedokończonych Zdań Rottera, Test Drzewa Kocha, Próbę Zdań Niedokończonych Żelazowskiej, Test Uzupełniania Zdań Sacksa i Levyeego, Serię Zdań Niedokończonych Hessa i Handela<sup>1</sup>, Test Zdań Niedokończonych wg Janusza Kostrzewskiego,

<sup>1</sup> Pochodzący z zapomnianej pracy Roberta Hessa i Geralda Handela [18]; polski tłumacz nieznan.

Test Niedokończone Opowiadanie Thomasa<sup>2</sup>, Test Rysunek Rodziny L. Cormana<sup>3</sup>, Test Film R. Gille'a<sup>4</sup>, Test Drzewo Przeżyć<sup>5</sup>, Test Kolorów Maxa Lüschera. Tabela I pokazuje, w jakim procencie spraw zostały zastosowane poszczególne techniki.

Ostaszewski pisze też [28, s. 205], iż bardziej obiektywne w jego opinii metody kwestionariuszowe są wykorzystywane zdecydowanie rzadziej niż projekcyjne. Wysoce niepokojący jest fakt wykorzystywania w tak ważnych sprawach technik o niejasnych podstawach i nieposiadających ewidencji empirycznej w tych dziedzinach. Można zaakceptować argument [28, s. 205], że trudno stosować techniki kwestionariuszowe wobec często 6-letnich dzieci. Jednak, jak pisze Ostaszewski [28, s. 188–189], wykorzystuje je się także z tego powodu, że podstawową ich zaletą jest ich „możliwość wyeliminowania celowego wprowadzania w błąd przez badanego (...) oraz to, że przedstawiają całościowy obraz psychicznego funkcjonowania badanego”. Paweł Ostaszewski opisuje też [28, s. 190] wyniki analiz Wandy Stojanowskiej, która w 2000 roku badała m.in. psychologiczne opinie RODK w sprawach o rozwód. W 25% akt wykorzystywano przy formułowaniu opinii testy psychologiczne, a wśród nich takie techniki, jak: Test Drzewa Kocha, Test Rodziny według M. Braun-Gałkowskiej, Test Postaci Ludzkiej Halpern i Machower<sup>6</sup>, Test Uzupełniania Zdań Sacksa i Sidneya<sup>7</sup>, Test Wolnych Skojareń Rottera, Kwestionariusz Zdań Niedokończonych dla Dzieci J. Kostrzewskiego, Test Barw M. Lüschera, Test Film R. Gille, Zdania Niedokończone wg Hessa, Zdania Niedokończone Żelazowskiej, Zdania Niedokończone dla Małżonków, Test projekcyjny „Dwa Domki”. W raporcie „Monitoring pracy wybranych Rodzinnych Ośrodków Diagnostyczno-Konsultacyjnych” prezentowanym na konferencji organizowanej grudniu 2009 roku w siedzibie Fundacji im. Stefana Batorego [13] wskazano, że w zanalizowanych 81 opiniach sporządzonych w latach 2002–2009 aż w 64% opinii wykorzystano różne metody projekcyjne o niejasnym statusie i mylących nazwach (m.in. Rysunek Rodziny Anny Frydrychowicz, Test Drzewa M. Kocha, Test Zdań Niedokończonych dla Dzieci Żelazowskiej, Test Rysowania Rodziny Cormana, Próbę Zdań Niedokończonych dla Dorosłych Sach-

sa i Levyego, Próbę Zdań Niedokończonych dla Dzieci Starszych Kostrzewskiego, Test Barw M. Lüschera [28, s. 37], w czym tylko metody projekcyjne wykorzystano jako bazę opinii w ok. 15% [28, s. 43]. Autorzy podają, że TD wykorzystano w ok. 28% zanalizowanych opinii [28, s. 42]. Nawet biorąc pod uwagę nielosowy, celowy dobór analizowanych opinii (w samym 2008 roku RODK wydały 23 045 opinii, w tym 9 247 w sprawach opiekuńczych [28, s. 14], niepokój musi budzić fakt, że w tych kontrowersyjnych opiniach aż w takim stopniu korzystano z technik projekcyjnych.

#### 4. O potrzebie wiarygodności narzędzi i o sposobach jej oceny

Wiarygodność oznacza, że coś jest godne zaufania, sprawdzone. Innymi słowy, psychologiczne narzędzia diagnostyczne to takie, na których użytkownik może polegać, bo ktoś przedtem sprawdził, czy są one miarodajne: czy są wystarczająco czułe, swoiste i efektywne [38, s. 50]. Testy psychologiczne powinny być wystandaryzowane, posiadać aktualne i właściwe dla celu badania normy (próby referencyjne), rzetelne i trafne. Powinny też być obiektywne – zarówno więc bezstronne (w sensie niezależności od diagnosty, nietendycyjne), jak i intersubiektywnie sprawdzalne oraz wolne od ocen (wartościowania).

W Standardach APA [1] opisujących podstawowe zasady stosowania testów nie wyróżnia jakiejś specjalnej kategorii „testów projekcyjnych”, których użycie miało by podlegać jakimś odrębnym regułom. Traktuje się je jak inne testy psychologiczne. Jedynie w komentarzu do standardu 12.12 napisano: „Gdy wykorzystuje się testy z nieustrukturyzowanym formatem odpowiedzi, takie jak techniki projekcyjne czy nieformalne oszacowania zachowania, to – gdy tylko to możliwe – należy przestrzegać jasnych i obiektywnych kryteriów punktacji, minimalizujących subiektywność ocen” [1, s. 228]. Nie ma więc powodu, by stosować do nich inne kryteria wiarygodności niż do pozostałych narzędzi diagnostycznych.

Przyjęto wiele kryteriów ustalania wiarygodności, głównie na rzecz opinii sądowych. Pierwsze z nich – tzw. kryteria Dauberta (ang. Daubert standard) – powstały w Stanach Zjednoczonych [30, 37, 38]. Pierwotnie chodziło o to, by zapewnić oskarżonemu poczucie bezpieczeństwa, że badania będące podstawą opinii (dowodu, ekspertyzy) wykonywane są według jasno określonych i sprawdzonych kryteriów. Inne kryteria odnoszące się już wprost do wiarygodności testów psychologicznych zaproponował Kirk Heilbrun [17]. Rozwinął je i uzupełnił, tworząc model mieszany naukowo-prawniczy, Douglas B. Marlowe [24]. W Polsce wiele uwagi poświęciła tej sprawie Alicja Czerederecka [11]. Podsumowując te

<sup>2</sup> Być może chodzi o test Madeleine Thomas [41]; polski tłumacz nieznan.

<sup>3</sup> Przepuszczalnie pochodzący z pracy Cormana [10]; polski tłumacz nieznan.

<sup>4</sup> Być może chodzi o test René Gille [14]; cytowany na s. 248 pracy Rembowski [33].

<sup>5</sup> Nie udało się ustalić źródła.

<sup>6</sup> Chodzi o test rysowania postaci ludzkiej stworzony przez Karen Machover, por. przypis następny.

<sup>7</sup> Chodzi oczywiście o test Josepha M. Sacksa i Sidneya Levyego [35]. Już ten fakt – nieznamość nazwisk autorów testów – pokazuje niski poziom autorów opinii, którzy się na wyniki tej techniki powołują.

propozycje, można je ująć w postaci następującej listy kryteriów:

1. Czy teoria (koncepcja) leżąca u podstaw testu jest sprawdzalna? Czy da się ją empirycznie weryfikować?
2. Czy teoria (koncepcja) leżąca u podstaw testu odnosi się do obszaru analizowanego w opinii, dla której test jest podstawą (wprost lub pośrednio) i czy było to przedmiotem badań jej trafności?
3. Czy i jakie są dowody empiryczne, że dana technika diagnostyczna jest logicznie powiązana z przedmiotem sprawy lub jest powszechnie stosowana w takich sprawach?
4. Czy dana technika diagnostyczna została opisana w recenzowanej publikacji w literaturze fachowej? Czy była przedmiotem środowiskowej dyskusji naukowej?
5. Czy dana technika diagnostyczna jest akceptowana przez przedstawicieli ośrodków naukowych zajmujących się daną problematyką? Czy jest ona zgodna z aktualnym stanem wiedzy w danej dziedzinie, lub ogólniej – czy jest zgodna z obecnym stanem nauki?
6. Czy dana technika diagnostyczna jest dostępna w sprzedaży wraz z podręcznikiem, odpowiednimi normami oraz znane są jej podstawowe parametry psychometryczne (w tym dla tego typu spraw/populacji)?
7. Czy zostały ustanowione standardy dla stosowania danej techniki diagnostycznej? Czy ma ona opisaną standardową procedurę badania i obliczania wyników oraz ich interpretacji?
8. Czy dana technika diagnostyczna jest stosowana do populacji i celu, do których jest przeznaczona?
9. Czy znany jest poziom błędów uzyskiwanych przy stosowaniu tej metody? Jaki jest błąd pomiaru techniki? Czy znany jest możliwy wpływ zakłóceń na wyniki danej techniki diagnostycznej oraz jej podatność na oszukiwanie?
10. Czy diagnosta posiada odpowiednią i udokumentowaną wiedzę i doświadczenie do posługiwania się daną techniką diagnostyczną?
11. Czy diagnosta uwzględnia probabilistyczny charakter wyrażanej opinii oraz czy opinia ma charakter konkluzyjny?

Powstaje pytanie, w jakim stopniu i które techniki projekcyjne spełniają te wymagania? Z pewnością nie jest tak, aby którakolwiek spełniała je wszystkie. Ale podobnie jest w odniesieniu do metody kwestionariuszowej. Wróćmy więc do pytania: co decyduje o wiarygodności narzędzia diagnostycznego? Wydaje się, że decyduje o niej ciągłe badanie trafności narzędzia. Powszechne jest błędne przekonanie, że tylko w fazie przygotowywania testu i jego komercjalizacji ważna jest poprawność metodologiczna procedur i – co za tym idzie – że po opubliko-

waniu test psychologiczny jest skończonym produktem (narzędziem). Jednak równie ważne jest to, co dzieje się po opublikowaniu narzędzia: czy następuje ustawiczna aktualizacja norm oraz badania naukowe nad obszarami jego trafności? Narzędzie diagnostyczne, na temat którego ukazują się doniesienia naukowe, jest zdecydowanie bardziej godnym wiary niż narzędzia, na temat których publikacji jest mało lub (poza podręcznikiem do testu) nie ma wcale. W latach 1940–2001 ukazało się [8] na temat testów MMPI i MMPI-2 blisko 14 tysięcy prac naukowych; rocznie publikowano średnio 226 ( $\pm 116$ ) prac. W latach 50. XX wieku ukazywało się rocznie średnio 111,6 prac na temat techniki Rorschacha i 26,1 na temat MMPI, w latach 1970–1979 ukazało się odpowiednio 64,6 prac na temat techniki Rorschacha i 115,4 prac na temat MMPI [32]. To ukazuje skalę problemu. Warto dodać, że na podstawie analizy cytowanej literatury w przeglądzie testów Burosa [7] widać wyraźnie, że od 1960 roku obserwuje się systematyczny spadek zainteresowania badaczy technikami projekcyjnymi na rzecz technik kwestionariuszowych<sup>8</sup>.

O wartości testu psychologicznego decyduje więc zarówno przestrzeganie standardów psychologicznych podczas jego tworzenia oraz podczas korzystania z niego w praktyce, jak i liczba prac naukowych, których jest on przedmiotem. Problem oparcia diagnozowania na dowodach naukowych oraz swoistej kontroli rynku komercyjnych narzędzi diagnostycznych jest ważny z jednej strony dlatego, że coraz silniejszy jest nacisk na analizowanie wiarygodności diagnostycznych narzędzi psychologicznych oraz dlatego, że – z drugiej strony – niewiedza potencjalnych klientów sprzyja swoistej „tabloidytacji” testów.

<sup>8</sup> Według wskazań programu „Publish or Perish” w latach 1940–2010 łączna liczba cytowań prac na temat techniki Rorschacha wynosiła 104 632, podczas gdy cytowania dotyczące MMPI/MMPI2 osiągnęły wartość 195 558.