

Testing of Intellectual Abilities in Various Sociocultural Environments

Denisa Denglerová

Contact to author

Masaryk University
Faculty of Education
Poříčí 538/31
603 00 Brno
denglerova@ped.muni.cz

Kontakt na autora

Masarykova univerzita
Pedagogická fakulta
Poříčí 538/31
603 00 Brno
denglerova@ped.muni.cz

Copyright © 2016 by the author
and publisher, TBU in Zlín.
This work is licensed under the
Creative Commons Attribution
International License (CC BY).



Abstract: This study concerns possibilities for fair testing of intellectual abilities in the context of different sociocultural environments. It traces the history of testing and misuse of test results to discriminate against various groups. Using examples of specific items from currently used WISC tests, the study shows how the environment influences test results. It then describes four models of the relationship between intelligence and tools to measure it in various sociocultural environments. In the last part of the study we propose a possible development of psychological testing that might make it truly fair and thus widely usable. We mention different paths to culturally appropriate tests with an emphasis on principles of dynamic testing.

Keywords: test of intelligence, cognitive abilities, fluid intelligence, culture relevant test, dynamic testing, socially disadvantaged environment

Testování intelektových schopností v různých sociokulturních prostředích

Abstrakt: Studie se zabývá možnostmi spravedlivého testování intelektových schopností v kontextu různých sociokulturních prostředí. Zabývá se dějinami testování a zneužitím testových výsledků k diskriminaci různých skupin. Na příkladech konkrétních položek z aktuálně používaných testů WISC ukazuje, jak prostředí ovlivňuje testový výsledek. Dále představuje čtyři modely vztahu mezi inteligencí a jejími měřicími nástroji v odlišných sociokulturních prostředích. V poslední části studie představujeme návrhy, jak by se mohlo psychologické testování vyvíjet, aby bylo opravdu spravedlivé a tudíž široce využitelné. Zmiňujeme různé cesty ke kulturně adekvátním testům s důrazem na principy dynamické diagnostiky.

Klíčová slova: test inteligence, kognitivní schopnosti, fluidní inteligence, kulturně relevantní test, dynamické testování, sociálně znevýhodněné prostředí

1 Introduction

The composition of our society is shifting from a relatively homogenous makeup to a more heterogeneous representation of members from various cultures. In psychology, taking into account the different needs of culturally diverse groups and the possibilities for meaningful mutual communication is the subject of intercultural psychology. Social pedagogy can benefit from the findings of intercultural psychology and transform its theoretical bases into meaningful interventions

in various sociocultural environments (whether these be immigrants, Romani or other socially disadvantaged groups). In the Czech environment, however, intercultural psychology emphasizes above all the subjects of differences among cultures in the areas of values, moral reasoning, attitudes to education, family education, gender stereotypes, communication styles, etc. (Průcha, 2010). Nevertheless, in countries with a greater number of immigrants and a long history of cultures mixing, intercultural, cognitive and developmental psychology in the last two decades has systematically examined how an individual's social and cultural background affects his or her way of thinking and understanding of the world (see, for example, Matsumoto & Juang, 2008; Kitayama & Cohen et al., 2010). Many studies show that cognitive abilities which traditional psychology had regarded as rather innate and little modifiable by the environment, are on the contrary strongly influenced by the environment, and their development is basically impossible to understand without taking into account the influence of the external environment (Denglerová, 2015).

This thesis about the cultural conditioning of cognition logically raises the question of how it is possible to fairly measure, evaluate and test cognitive abilities. This question, moreover, is not merely an academic exercise, because minority groups are tested by mainstream diagnostic tools in the Czech practice, and on the basis of thus poorly executed test examinations, or the interpretation thereof, the fate of specific individuals is decided (mostly in preschool diagnostics and diagnostics of school readiness and classifying children into a specific type of educational establishment, and furthermore in e.g. selection procedures for work positions). In our country, this issue concerns mostly Romani children and adults from socially disadvantaged and/or culturally different environments. In cases when Romani children with adequate intellectual abilities are, at the beginning of their educational path, already included in practical schools, primarily intended for the education of mentally handicapped children, their education possibilities and later professional lives are restricted. Such segregation of children is not only unethical and against the Schools Act, it also amplifies the poor social situation in which a large part of the Romani community lives, transferring it on to the next generation (Human Rights League, 2005). Romani children's unequal chances to receive a quality education are, indeed, not caused merely by improper testing of their intellectual potential. There is an entire system of causes, from the content of school curricula and teaching methods that do not reflect the needs of Romani children, through the unreadiness of teachers to educate these children, to the low level of aspiration of the children themselves (Dočkal et al., 2004). Similar causes, including inadequate assessment of cognitive abilities and ordinary teachers' lack of interest in the minds of Romani children, are also reported by Portik (2003).

In this theoretical study, however, we deal with the options of fair testing across¹ varying sociocultural environments. We use the history of testing to demonstrate the discrimination that results from inadequate testing or misinterpretation of test results. This discrimination emerged shortly after the first tests were made available. For example, the immigration policy in the United States in the 1920s was made significantly stricter on the basis of the low results that European immigrants scored in tests of intelligence (Gould, 1998). After a brief look at the history of the testing of intelligence, we focus on the relationship between intelligence and the measurement of it in the contemporary humanities. Then, we outline possible ways for taking into account sociocultural aspects of the environment when testing intellectual abilities, so that no tested proband is discriminated against.

The most discussed, and most tested area of cognitive abilities is intelligence. This is somewhat ironic, as despite the great attention paid to intelligence in the fields of pedagogy and psychology,

¹ When the term "testing" is used throughout this text, it is meant as testing of cognitive abilities, especially intelligence. We are not primarily concerned with other personal characteristics, although the majority of ideas on how to correctly use a test and not discriminate against any population group on the basis of it, are also applicable to general principles of testing. The history of testing over its two decades also equals the history of testing of the cognitive abilities. It was not until 1921 that Hermann Rorschach published his complex test, thus shifting attention from testing intellectual abilities to testing the personality as a whole.

experts do not agree on its exact definition. Various definitions differ primarily in what they include in intelligence. This study does not endeavour to provide an exhaustive comparison of various approaches to intelligence. At first, we will need only a general understanding of intelligence as a disposition to solve problems and an ability to adapt to conditions imposed by the environment. We will then further refine the definition of intelligence, and modify it according to the individual authors of intelligence tests and their own interpretation. Given that the text analyses options and methods of measuring intelligence (including their erroneous or improper implementation) we implicitly assume that intelligence is measurable. In defining this ability we assume the psychometric attitude towards intelligence, which stresses the measurability of intelligence. Even though there are hundreds of definitions of intelligence, most of them include the reaction of an organism to its external environment and emphasize adaptation to this environment as a key component of intelligence. The environment and its impact on the shaping of a personality, as well as working with the environment in which a man develops, are the remit of social pedagogy. Bakošová (2005) reports that social pedagogy can be understood as learning by environment, and education within its framework is understood as a help for all age categories, including the preschool period. If we wish to increase the test success rates of people from socially disadvantaged environments, we need to be aware of their environment's specific features, we must evaluate and appreciate their ability to adapt to their primary environment and not the environment of the majority population.

2 History of the testing of intellectual abilities

The dawn of testing of intellectual abilities was quickly followed by misuse of tests to exclude people. Alfred Binet² is considered the first creator of intelligence tests as we know them today. Originally he had no intention to develop such tests. In accordance with the dogma of his time, he performed craniometric³ research, but he repeatedly discovered that the skull circumferences of children that he measured had no connection to the intellectual performance these children displayed at school. Binet bravely published these findings, despite contemporary scientific authorities being convinced otherwise. In 1905 he published the first version of the intelligence scale, which was supposed to detect children who, due to their lower intellectual abilities, would perform poorly in the classic school lessons, and therefore should be given special pedagogical care (Svoboda, 2010). The test was created at the order of the French Ministry of Education. Interesting for us is Binet's attitude to testing of children's intellectual abilities. He emphasized that his tests measure the "state of the intellectual capacity of a child at the given time, nothing more" and he focused on "purely and simply establishing the truth regarding the current mental state of the child" (Gould, 1998). By this he endeavoured to emphasize that his test would not capture the developmental aspects of intellectual abilities, nor would it predict the possibility or impossibility of a future change. The intellectual deficits revealed by Binet might have various causes and he himself did not dare to assess whether they were inherited or acquired during development. He was moreover convinced that when given appropriate assistance, each child could improve, although some more than others. According to Binet, the special interventions for children with intellectual deficiency cannot be the same for every child, and it is crucial to make accommodation for the character, abilities, needs and potential of each pupil.

As a first step he recommended a substantial change of environment and education in small classes of about 15 children (compared to the usual 60 children in France of the beginning of the 20th century) (Gould, 1998). Although these ideas are a century old now, they correspond with our

² Alfred Binet was a French psychologist, born in 1857 and working in the late 19th and early 20th century.

³ Craniometry is an anthropological method used to measure the size and shape of the skull. Until the beginning of the 20th century, there was a strong conviction in Europe and the USA that the size of the skull correlated with intelligence.

view on helping children who score low in intelligence tests.⁴ However, the ideas seem to have been forgotten during the 20th century and an understanding prevailed of intelligence as an unchangeable feature that was little subject to influence. Measuring of intelligence thus became a very important act that segregated children based on the single-time readings obtained of their intellectual abilities, predetermining the course of their education (and thus, to some extent, also of their lives).

The particular way of using intelligence tests and interpretation of their results was thus connected with abuse of the tool, whether intentionally or as the result of inappropriate application. Binet's original thoughts were abandoned quite early: When his test spread over the USA it was adopted by supporters of hereditarism,⁵ who started to use its translated and modified versions to validate their opinions. They thus did not respect Binet's basic thesis and considered the test outcomes unchangeable, characterising the intellectual abilities of test subjects not only in the present, but also in the future. When carrying out tests on children and adults, the hereditarists suggested separating individuals in which intellectual deficiencies were detected from society, rather than offering them assistance in developing their abilities. The tests were often performed in unsuitable and exhausting conditions. An extreme example of such completely inappropriate testing was the examination of the intellectual abilities of the immigrants right after they landed at Ellis Island, after the long sea voyage from Europe. By this testing, Goddard succeeded in "proving" that among immigrants, many intellectually deficient people were entering the USA, and thus successfully increased the number of people denied access to American soil based on poor test results (Gould, 1998). Such obvious examples of the abuse of intelligence tests aside, we still find other errors in their usage. Another shift from Binet's intentions was the use of his scale to evaluate the intellectual abilities of intact children,⁶ and subsequently comparing results and making recommendations as to what education level children should receive.

The above-mentioned scientists who tested intellectual abilities were building on empirical experience. They did not rely on any theory that would describe intellectual abilities in detail. Binet himself pointed this out and he considered the test to be a mere simple tool to meet the order of the Ministry of Education. He probably would have been quite surprised to see how the tests were employed and what importance they gained in the examination of human subjects.

The first factor-analytic model was developed in the 1920s by Charles Spearman, who with the help of factor analysis concluded that intelligence consists of a general factor g and specific factors (Ruisel, 2001). The general factor pervades the entire intellectual performance of a person. Spearman also studied what determined the level of the general factor g . He came up with a term "mental energy," including (without further specification) the subject's attention, the plasticity and complexity of the nervous system, etc. (Sternberg, Kaufman, & Grigorenko, 2008). Spearman did not directly address the question of whether intelligence is innate or influenced by the environment, but the qualities he hides under the concept of mental energy rather refer to biological potential and therefore a greater genetic determination of intelligence.

Cattell disagreed with Spearman's concept of intelligence as a uniform ability and already in the 1940s he divided Spearman's g -factor into two separate factors. Cattell considered "fluid intelligence" the general ability to assess and understand relationships, and he assumed that this ability is largely genetically determined. He defined "crystallized intelligence" as an ability that develops due to learning and life experience, and so education may have a positive effect on its growth, but a sufficient level of fluid intelligence is a prerequisite for the development of crystallized intelligence (Plhánková, 1999). In its time, this classification was quite groundbreaking and especially the thesis that a person's environment and education are responsible for a particular part of his or

⁴ We purposely leave aside the question of inclusion in education for the time being, focusing on the idea of supporting children with low scores in intelligence tests.

⁵ Goddard in the 1890s and Terman in the 1920s represented leading figures of hereditarism.

⁶ They used Binet tests as a sieve to capture children with low intellectual performance.

her intelligence was something new. From a modern perspective, we must say that Cattell still underestimated the influence of the environment. Cattell focused primarily on testing of fluid intelligence and he developed a test which, he believed, measured the intellectual abilities of a person regardless of his or her sociocultural environment. In 1949 he published the first version of the Cattell Culture Fair Intelligence Test, which was, according to its author, suitable for testing the intelligence of children from eight years of age and adults, and it contained items that focus on deducing relationships between graphic symbols (Fajmonová, Hönigová, Urbánek, & Širůček 2015). This test has been particularly beneficial by filtering knowledge from school, otherwise culturally determined knowledge, and knowledge gained through the course of one's life out of the concept of intelligence. However, Cattell's apparently well-intentioned effort to design a culturally fair test was mainly focused on removing the verbal content of the test (the instructions, however, were given verbally) and thus minimizing the influence of one's mother tongue. The test could thus be useful for testing people growing up in a given (American) culture, whose native language is not English (but a certain level of language knowledge was expected and necessary for understanding the test instructions).

Even different ways of thinking and their preferences (in the case of the Cattello test, searching for relationships between objects) are, however, culturally conditioned. For instance, Nisbett (2003) through his repeated studies has shown differing preferences of representatives of Western and Eastern cultures in the spontaneous sorting of objects. Other researches have shown the influence of cultural context on fluid intelligence as well, for example by comparing the overall cognitive development of children from disadvantaged minorities with that of children growing up in stimulating environments. Cattello's concept of intelligence, revolutionary in its time, is still widely cited and used (even abused) by supporters of hereditarism to enforce the proposition that most of the differences in cognitive abilities between people or various groups of people are determined by heredity, and therefore the effort to develop what little can be influenced by upbringing and education is ineffective and uneconomical for society as a whole. Hereditarists are convinced that there simply are groups of people, usually ethnically or socially defined, with innate lower intelligence ability (Denglerová, 2015).

A further advance in testing the development of intelligence was brought by David Wechsler's tests. The first of these, designed for adults, was published in 1939. Both the versions for adults and the versions for children that were developed later have been revised and modified numerous times since then. Wechsler perceived intelligence as the complex and global ability of an individual to act, think rationally, successfully cope with its environment, understand the world and effectively handle its challenges (Krejčířová, Boschek, & Dan, 2002). This definition has two fascinating aspects: How modern it actually is, and how it emphasizes the aspect of the environment a person lives in. However, because of its connection with the environment that surrounds a person, the test may not be suitable for individuals outside the mainstream of society, and this quite logical reservation often goes unnoticed by the professional community.

Each version of Wechsler's test consists of two basic parts, a verbal one and a non-verbal one. In a simplified way, these parts correspond to Cattello's classification of crystallized and fluid intelligence. The verbal part of the test includes items that test mainly the ability to understand and use language, vocabulary, to logically extrapolate from verbally specified content, and the like. These skills are generally considered to be subject to influence from learning and the surrounding environment. The non-verbal or opinion section consists of items which are given mainly in the form of images and symbols, and it is believed that it tests in various ways fluid intelligence, and therefore it is little influenced by the environment shaping a human being.

This concept is, however, erroneous, as we illustrate by several of the items from the currently used version of Wechsler's tests⁷ for children. For example, in one of the subtests the child is instructed to sort flashcards which, in a correct order, form a certain short story. Depending on the task difficulty, there are 3 to 5 cards. One of the simpler items in this test contains three images (listed in the correct order): A child climbs a ladder to a slide, the child sits on top of the slide, the child slides down the slide. If the tested child knows the principle of a slide, he or she has no problems sorting the cards accordingly. However, the item certainly does not test anything that would be connected to the ability to learn new things or understand the world and its challenges. It tests the child's knowledge, or rather his life experience based on the thousands of various slides that he or she has climbed up on and slid down (or watched this being done by other children on the playground). On the other hand, for a child who has never been to a playground and has not gone down a slide, this picture can actually serve as an impulse for thinking and understanding the principle behind the picture. A smart child can come up with a correct order for the pictures without actually being familiar with any slide (the item could therefore be more valid for children who have no experience with slides), but it will take some time. Sorting pictures in Wechsler's test is not rated only as success or failure, bonus points are given for quick sorting of the picture series. Such a rating clearly discriminates against children without the experience illustrated on the pictures, and as such is strongly dependent on the sociocultural environment of the children (Denglerová, 2015).

In another type of task, a child was supposed to reveal the relationships between the objects pictured. The tested child is, for example, shown a picture of an open refrigerator and a bottle of milk. On another picture, the child sees several pieces of folded laundry and is instructed to choose a picture that is in a same relationship to this picture as the bottle of milk to the fridge. The child can select from a mailbox, a stove, an open suitcase and a handbag. A child with no experience with a refrigerator and no knowledge of it being used for food storage cannot replicate the same relationship, so this item again invisibly depends on a certain knowledge (however common such knowledge is in the mainstream society), and therefore it is not suitable for testing children from a significantly different cultural background (in this case children from poor families living in socially excluded areas without the usual appliances). Nonetheless, items not depicting particular things and everyday objects are not a solution either. If a child is, for example, supposed to find a symbol to fill in an empty space in a row of symbols (a verified way of evaluating abstract thinking) and has never done a similar task before, nor has he or she ever been in a situation that required this way of thinking, this child will probably not be able to solve the problem correctly, or the child will need a substantially longer time to solve it than his or her peers who have more experience with a similar way of thinking.

3 Models of the relationship between intelligence and tools to measure it in various sociocultural environments

Current views on intelligence in a sociocultural context and the possibilities for measuring it are reviewed by Sternberg (2004). He does not address a specific definition or particular concept of intelligence, but he assumes that in all cultures it is the ability to solve new problems and adapt to new situations. He focuses on two basic questions, namely whether there are innate differences in the mental processes contained in intelligence across cultures, and whether there are fundamental differences among tests measuring intelligence, arising from cultural differences (Sternberg, 2004).

⁷ WISC III is the most-used complex test for testing intelligence in children aged 6–16 in the Czech Republic. It was first published in the Czech Republic in 1996, and again in 2002 (Krejčířová, Boschek, & Dan, 2002), this time with Czech standards. Before that, a PDW variant was used, and WISC IV is the version used most around the world today.

Based on the answers to these questions and their combination, Sternberg introduces four models on how to relate to intelligence across various cultural environments.⁸

In the first model, the nature of intelligence is completely the same across all cultures, and therefore the principles of evaluating it should be the same. According to this opinion, intelligence across cultures should be measured by the same intelligence tests. The tests should indeed be appropriately localized in terms of language, but apart from that there should be no modifications for any particular culture. We fundamentally disagree with this model; the arguments against it follow from the above-mentioned analysis of items in these tests of intelligence. A further proof of the inaccuracy of the model is also the generally accepted hypothesis of linguistic relativity, according to which the human understanding of the world is influenced by the language skills of a particular society (Budil, 2003).

In the second model, we encounter a different concept for intelligence, but the same measurement tools. Intelligence tests or their individual items, however well translated, thus measure something structurally different in each culture. For example, Nisbett (2010) used the same tests to show how differently people think about things in different cultures. This approach has its relevance, but comparisons of such results across cultures or even their evaluation in terms of "better versus worse" must be strictly rejected.

The third model assumes that the nature and dimensions of intelligence are the same in all sociocultural environments, but it is necessary to measure them by different tests adequate to each culture. This does not mean that no intelligence tests could be used for measurements across different cultures, but besides a high-quality translation, we must not forget about the emic aspect and ensure that the psychological meaning of the test question remains, or modify the items so that the original meaning is transferred to another culture (Denglerová, 2015). Sternberg (2004) considers this third model the most suitable and he claims that the concept of successful intelligence that he had formulated before is in agreement with this model. This model is widely used (more or less reflectively) also among researchers dealing with minorities living in socially disadvantaged environments.

This can be illustrated by the example of an item from the intelligence⁹ test developed and validated on children from Romani settlements in eastern Slovakia by Dočkal and others (2004). In the picture task, children are instructed to arrange four flashcards in a logical order. The pictures display a girl waking up, getting dressed, having breakfast and leaving to school. To make the pictures understandable for children from Romani settlements, some objects that were unknown and confusing had to be left out, among others a clock above the bed, a carpet in the room, a school bag, etc. Even after that, the children kept arranging the pictures in a different order than expected. A common sequence was waking up–getting dressed–leaving to school–having breakfast (compared to the expected sequence of waking up–getting dressed–having breakfast–leaving to school), because most of the children from the settlements do not get any breakfast and they are given the first meal of the day at school. This item, therefore, had multiple correct solutions among children from poor Romani settlements, which is not common for classic intelligence tests. Yet when it was modified and adapted to the conditions where the children were growing up, the item distinguished well between

⁸ In a similar way we could view and combine models also for other characteristics beside intelligence, and their measurement across cultures. Indeed, nobody in the scientific community today would consider, for example, emotions and the way of expressing them to be the same in all cultures, and nobody would attempt to evaluate such using the same key.

⁹ This is a battery of tests published as an RR screening, which is designed for children aged six to ten years old, and the aim of which is to distinguish children with mental retardation from children with intact intellect. The test was custom designed to re-measure the intellectual abilities of Romani children in Slovakia attending special schools (similar to Czech practical schools). Based on the test results, efforts are being made to integrate children for whom mental retardation has been ruled out into mainstream schools and thus increase their level of education (Dočkal et al., 2004).

children with mental retardation (who arranged the flashcards in a random order and failed to explain their choice) and children with intact intelligence (Dočkal, 2007). In a similar way, other items of the test were also modified.

The fourth model assumes that the nature of intelligence and its various dimensions are highly dependent on the culture an individual is a part of, and also the measuring instruments must be different. This approach is radically constructivist and emphasizes that intelligence can be understood and measured only in the context of the original culture (Denglerová, 2015). This approach can be illustrated, for example, by the fact that Dočkal and his colleagues completely eliminated from their test battery subtests aimed at classification and elimination, i.e., tasks that require the ability to generalize. They claim that such mental operation is hardly developed in children from poor Romani settlements, because they do not need it in their environment, and therefore it makes no sense to test for it within tests of intellectual abilities (Dočkal, 2007). In contrast, in conventional intelligence tests designed for the mainstream population, items utilizing the principle of generalization are widespread and well discriminatory. The question remains whether, in this constructivist conception, a researcher coming from the outside (and therefore a different cultural framework) can actually capture all the nuances and subtleties of the cognitive skills of the studied population and create an adequate measuring tool.

Based on our own experience and research, we do not dare to clearly answer Sternberg's question of whether there are innate differences across cultures in intellectual processes. We are aware of a significant impact of the sociocultural environment on the development of cognitive skills, as we have reported in previous studies (Denglerová, 2012b, 2015). At the same time, however, we do not deny the possibility that there are some innate aspects of cognition. On Sternberg's second question regarding the possibility or even necessity of differences between tools to measure intelligence in different cultures, we clearly answer that it is necessary to adequately modify and adapt the tests to match the structure of thinking that is characteristic for the particular culture. In our reflections, therefore, we do not consider the first and second model meaningful, and we work within the third and fourth models.

4 Potential ways of non-discriminatory testing

The effort to develop an intelligence test which would neither discriminate against nor treat preferentially any group of people is more than half a century old. It goes back to Cattello's attempt to create a culturally independent test as we described earlier. Today's perspective, however, shows that it is almost impossible to develop a culture-independent test, because every person is strongly shaped by his or her culture and is basically an inseparable part of it. Items that would constitute such a test would have to be aimed at testing the ability of abstract thinking (because with such items we could at least partially abstract from cultural determinants). Such items would also have to be presented to tested persons after thorough training and explanation of the principles behind the items, which would be closer to the principles of *dynamic diagnostics*. Additionally, a test with items focused only on the level of abstract thinking would not be relevant for the majority understanding of intelligence as an ability to navigate effectively in the real world, and it would only evaluate a small slice of the capabilities included in the concept of intelligence (Denglerová, 2015).

The next step in the development of non-discriminatory tests of intelligence was represented by so-called culturally fair tests with verbal items that relate only to those objects and events that should be known to all people regardless of their social background. The idea of such tests is based primarily on a contextual approach to intelligence. But even here many authors have come to the conclusion that it is not possible to construct a culturally fair test, given that performance is nearly always influenced by cultural factors (Konečná, 2010). In recent years, the term "culturally relevant intelligence tests" (Sternberg, 2001) has appeared, taking into account the social background of the person studied. Synonymous with "culturally relevant" is the term "culturally adequate," which was

probably first used in the DSM-5 (Raboch et al., 2015), where the part on diagnosis of various mental illnesses, including cognitive deficits, emphasizes that tests and methods that are adequate to the sociocultural environment from which the diagnosed person comes from must always be used for the final diagnosis. Standards for pedagogical and psychological testing (Klimusová, 2001) are also based on the principle of fair attitude to testing, and rule out – among others – items which “...lead to a different meaning of scores gained by members of different categories of people...”

In the last part of the study, we introduce suggestions for how tests of cognitive abilities could be developed to be truly fair and thus widely usable. We suggest and analyse various ways to create culturally adequate tests. A relatively simple, albeit tedious, way is to build on existing intelligence tests, which should, however, undergo a new process of standardization that would focus on specific populations, primarily on children from Romani families. It is necessary to concentrate on a detailed analysis of items and decide which ones are valid in this population considering the intended characteristic tested. Some items just need to be modified, while others would have to be completely excluded from the test. An example of such analysis of individual items is Dočkal’s RR screening (Dočkal et al., 2004) as mentioned above.

Another way to avoid the use of conventional tests of intelligence (and thus their potential cultural conditioning) is measuring of the so-called *adaptive abilities* of an individual. Adaptive abilities describe how individuals are able to function (adequately to their age) in their sociocultural environment and to cope with the demands of everyday life. Adaptive skills are assessed in the natural environment of the individual with respect to all aspects of his or her life. Evaluation of adaptive capabilities is usually employed to search for children and adults with mental retardation and does not aim to measure and compare the intelligence of people within broader standards.

We prefer the option of dynamic testing, which stresses the process of learning in cognitive abilities and attempts to evaluate it. Dynamic diagnostics is primarily focused on the extent and nature of change that will take place in the test subject after the application of a certain intervention strategy, so it is a test of learning ability and adaptation, in the broadest sense of the term (Chuchutová, 2008). The concrete possibilities for the technical implementation of dynamic testing are several, but the very essence always remains – trying to help the tested person to improve their score. That means that during intervention we provide feedback on the individual’s performance, help him or her clarify the principle behind the items, explain any mistakes made, and so on.¹⁰ Only the second or subsequent testing will show whether and how the tested person is able to improve their performance. The intervention may be minimal, lasting several minutes, or may take the form of several hours of training. For example, Grigorenko (Sternberg & Grigorenko, 2002) reports on her experiences with examining various cognitive abilities in children from rural areas in Tanzania. It was discovered that for a child to substantially and permanently improve his or her performance in deducting conclusions from linear syllogisms,¹¹ a five to ten-minute intervention clarifying the principle of the tasks in verbal and graphic form was usually sufficient. Also in our research, primarily focused on visual perception (Denglerová, 2012b), Romani children from socially disadvantaged environments far more often failed in distinguishing between pairs of identical and different elements. After about five minutes of intervention explaining the terms “same” and “different” and demonstrating these by means of wooden dice, most children markedly improved their performance.

On the contrary, in one of the sophisticated systems for diagnostic testing, the so-called LPAD (Learning Potential Assessment Device) by Reuven Feuerstein, interventions aimed at development of the tested child occupy most of the time allotted to the examination of each child. Such

¹⁰ This is in direct contradiction with the procedures of classic static testing. Providing feedback other than information about the achieved score, e.g. error analysis, is considered as a source of measurement errors (the training effect), and therefore is not carried out.

¹¹ Children worked with items such as “Adam is taller than Albert. Cyril is taller than Adam. Who is the tallest?”

examination usually lasts from several days to a week. As early as the 1950s, Feuerstein argued against the majority belief of educators and psychologists that intelligence is innate and immutable. He claimed that intelligence is not something that is given to us once and for all, but on the contrary it is a dynamic quality of a person, which can be developed in any stage of life (Feuerstein, Feuerstein, & Falik, 2010). This represents the base of his test battery. It is a set of tests on the development of the ability to learn. Some items are directly taken from Raven's matrices or Rey-Osterrieth's figure, while others are loosely inspired by them. However, testing involves an intervention part, which is focused on introducing and explaining a certain intellectual principle, and only then a successive solving of other items, based on a similar principle in various forms and difficulty levels, is evaluated. Experts agree that compared with traditional diagnostic methods, dynamic testing provides a lower level of discrimination, and therefore it is suitable for ethnic minorities, socially disadvantaged people, and children with learning disorders or affective disorders (Chuchutová, 2008).

It is particularly suitable for measuring the cognitive abilities of children from socially disadvantaged environments, since no emphasis is given to the element of surprise and to seeking out what the child in his or her cultural environment has not encountered, but rather makes it possible to eliminate these shortcomings within the intervention and to identify the child's potential during the learning process. A disadvantage of dynamic testing, which has hindered a significant expansion of it, is its duration and therefore the economic costs. Despite this, we are confident that in the near future dynamic testing will be encountered more often, because it gives a sense to assessing the abilities of persons outside the mainstream population.

5 Conclusion

The aim of the article was to point out that the commonly used intelligence tests are not usually able to capture and process the ability of people outside the mainstream population. Their use in the testing of children from socially disadvantaged environments (in our context typically Romani children) leads to incorrect conclusions about these children's abilities. A look at this history of testing makes it clear that discrimination against various groups of the population, mostly immigrants or the poor, occurred soon after the first tests emerged. In this text we have attempted to give solid examples of the most common erroneous principles on which the intellectual abilities of people from socially disadvantaged environments are evaluated, without clearly declaring that the test records rather poverty or a difference of environment and a potential impact of this environment on the development of intellectual abilities.

The study also highlights the principle of dynamic diagnostics that is not widespread in the Czech Republic, but which still has a potential to overcome the problems behind the classic cognitive testing. In terms of capturing the impact of the sociocultural environment on the development of cognitive skills, we consider dynamic testing to be a suitable alternative.

Social pedagogues are not expected to be able on their own to construct new culturally relevant and adequate tests of intellectual abilities, but should have a view on the issue and take the results of conventional tests of intelligence presented by their clients with reservations. We hope that in the near future, social pedagogues will be a part of the teams creating the new tests, and they will ensure consideration of environmental factors. Also within dynamic diagnosis, social pedagogues should be trained to provide appropriate interventions for persons tested that come from a different sociocultural environment.

We also want to point out that the naming of a particular behaviour, where children from disadvantaged backgrounds display lower performance, matters. For example, Ramey and Ramey (2000) show how negative naming of issues affects the direction of their solutions. Repeated studies in the USA and the UK have showed that children exhibiting the same or similar problems in school

were significantly more often diagnosed as “mentally retarded” if they came from the lower social classes, whereas if they came from a middle class, they were usually labelled as “having problems or learning difficulties.” Children with learning problems and disorders are given more attention and expert intervention in an effort to help them, while such assistance for the children labelled as mentally retarded is severely limited. Also, inappropriate use of intelligence tests may further contribute to referring to children outside the mainstream population as mentally retarded and thus significantly reduce their chances of a meaningful education appropriate to their needs.

This study was supported by grant no. GPP407/11/P091 from the Grant Agency of the Czech Republic within financing the project “Analysis of Knowledge Spaces by Children from Socially Disadvantaged Environment.”

References

- Bakošová, Z. (2005). *Sociálna pedagogika ako životná pomoc*. Bratislava: Univerzita Komenského.
- Budil, I. T. (2003). *Mýtus, jazyk a kulturní antropologie*. Praha: Triton.
- Denglerová, D. (2012a). Kognitivní schopnosti a jejich interkulturní souvislosti. In J. Němec (Ed.), *Výzkum a edukace sociálně znevýhodněných žáků. Romští žáci, děti cizinců (uprchlíků), teorie, výzkum, edukační strategie* [Research and education of socially disadvantaged pupils. Roma pupils, refugee children, theory, research, educational strategies] (pp. 71–82). Brno: Paido.
- Denglerová, D. (2012b). Percepce jako složka inteligence – Sociálně znevýhodněné děti na začátku školní docházky. In L. Gulová (Ed.), *Aktivizace sociálně znevýhodněných skupin v pedagogické praxi a výzkumu* [Activisation of Socially Disadvantaged Groups in the Practice and Research in Education] (pp. 106–112). Brno: MUNI Press.
- Denglerová, D. (2015). *Testování kognitivních schopností u dětí v kulturně rozmanité společnosti* [Testing of cognitive abilities of children in culturally diverse society]. Brno: MUNI Press.
- Dočkal, V. (2007). *Meranie schopností v interkultúrnom kontexte*. In L. Kolman (Ed.), *Komunikace mezi kulturami* (pp. 51–68). Praha: Alfa.
- Dočkal, V. (Ed.). (2004). *RR screening. Testová batéria na vylúčenie mentálnej retardácie 6–10 ročných detí*. Bratislava: ECO.
- Fajmonová, V., Hönigová, S., Urbánek, T., & Širůček, J. (2015). *CFT 20-R. Cattellův test fluidní inteligence* [CFT 20-R. Cattell culture fair intelligence test]. Praha: Hogrefe–Testcentrum.
- Feuerstein, R., Feuerstein, R. S., & Falik, L. H. (2010). *Beyond smarter. Mediated learning and the brain's capacity for change*. New York: Teachers College Press.
- Gould, S. J. (1998). *Jak neměřit člověka. Pravda a předsudky v dějinách hodnocení lidské inteligence*. Praha: Nakladatelství Lidové noviny.
- Human Rights League (2005). *Segregovaná výuka Romů ve školách střední a východní Evropy*. Praha: Artron.
- Chuchutová, K. (2008). Dynamické hodnocení a jeho využití u předškolních dětí. *e-Psychologie – Elektronický časopis ČMPS*, 2(2), 1–11. Retrieved from <http://e-psycholog.eu/pdf/chuchutova.pdf>
- Kitayama, S., & Cohen, D. (Eds.). (2010). *Handbook of cultural psychology*. New York: The Guilford Press.
- Klimusová, H. (2001). *Standardy pro pedagogické a psychologické testování (APA)* [Standards for educational and psychological testing]. Praha: Hogrefe–Testcentrum.

- Konečná, V. (2010). *Sebepojetí a sebehodnocení rozumově nadaných dětí*. Brno: Masarykova univerzita.
- Krejčířová, D., Boschek, P., & Dan, J. (2002). *WISC III. Wechslerova inteligenční škála pro děti – 3. vydání* [Wechsler intelligence scale for children – Third Edition]. Praha: Testcentrum.
- Matsumoto, D., & Juang, L. (2008). *Culture & psychology*. Wadsworth: Cengage Learning.
- Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently...and why*. New York: Free Press.
- Nisbett, R. E. (2010). *Intelligence and how to get it*. London, New York: W. W. Norton & Company.
- Plháková, A. (1999). *Přístupy ke studiu inteligence*. Olomouc: UP.
- Portik, M. (2003). *Determinanty edukácie rómskych žiakov*. Prešov: Pedagogická fakulta Prešovskej univerzity.
- Průcha, J. (2010). *Interkulturní psychologie. Sociopsychologické zkoumání kultur, etnik, ras a národů*. Praha: Portál.
- Raboch, J., Hrdlička, M., Mohr, P., Pavlovský, P., & Ptáček, R. (2015) *DSM-5. Diagnostický a statistický manuál duševních poruch* [DSM-5. Diagnostic and statistical manual of mental disorders]. Praha: Hogrefe–Testcentrum.
- Ramey, C. T., & Ramey, S. L. (2000). Intelligence and public policy. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 534–548). Cambridge: Cambridge University Press.
- Ruisel, I. (2001). *Základy psychologie inteligence*. Praha: Portál.
- Sternberg, R. J. (2001). *Úspěšná inteligence*. Praha: Grada.
- Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, 59(5), 325–338. doi:10.1037/0003-066X.59.5.325
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Sternberg, R. J., Kaufman, J. C., & Grigorenko, E. L. (2008). *Applied intelligence*. Cambridge: Cambridge University Press.
- Svoboda, M. (2010). *Psychologická diagnostika dospělých*. Praha: Portál.