The Psychological Assessment of Students with Diverse Developmental Needs – Universal Design Approach

Beata Papuda-Dolińska[1], Tomasz Knopik[2], Grażyna Krasowicz-Kupis[2], Katarzyna Wiejak[1]

[1] Educational Research Institute, Poland

[2] Institute of Psychology, University of Maria Curie-Skłodowska, Poland

HOW TO CITE:
Papuda-Dolińska, B., Knopik, T., Krasowicz-Kupis, G., & Wiejak, K. (2023). The Psychological Assessment of Students with Diverse Developmental Needs – Universal Design Approach. *International Journal of Special Education, 38*(3), 47-59.

CORRESPONDING AUTHOR:
Beata Papuda-Dolińska;
b.dolinska@ibe.edu.pl

DOI:
https://doi.org/10.52291/ijse.2023.38.38

COPYRIGHT STATEMENT:
Copyright: © 2022 Authors. Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:

The purpose of this article is to present the application of the universal design model in the broad field of psychological assessment for education, covering the area of creating diagnostic tools as well as designing the process of assessment. A universally designed diagnostic tool enables fair and valid assessments of a wide range of users, including individuals with special educational needs. It also helps prevent bias in test scoring and interpretation and, at the same time, enables fairness in test use.

The authors present the synthesis of the guidelines concerning the design of universal diagnostic tools and the diagnostic process itself at four key levels: conceptual, formal, test administration, and interpretative. As an example of a universally designed tool, the Comprehensive Analysis of Cognitive Processes (KAPP) has been mentioned. Discussion includes the benefits and limitations of applying the idea of universal design to psychological assessment with an emphasis on test development.

Keywords: universal design, universal design for assessment (UDA), psychological assessment, test construction, test fairness

INTRODUCTION

The idea of *Universal Design* mainly explains the principles of creating a physical space accessible to all its users, considering the needs of people with disabilities without the need for special adaptations (Mace, 1985; Nave, 2021). In terms of its meaning beyond the area of specific products and services, this concept has become a model guideline for complex and multifaceted phenomena such as education in its inclusive dimension. The original definition of universal design was adapted to the conditions of education (*Universal Learning Design*) and it has become a general model focused on the goal of creating a flexible learning environment, corresponding to the different educational needs of diverse groups of students (Rose & Meyer, 2002; CAST, 2008). In practice, this means preparing curricula, materials, and the environment in such a way that they can be appropriately and easily used by all the students in the teaching process (Bowe, 2000) over its entire span, including the educational objectives, content, methods, and aids. Compatible with the approach of universal design in learning is the idea of *Universally Designed Assessments* (UDA) for the assessment of school skills used in *large-scale assessments* (Thompson et al., 2002). Empiricism in this field confirms that both non-disabled students and students with disabilities obtain better results in universally designed tests than in traditional ones (Johnstone, 2003; Dolan et al., 2005). The UDA approach reduces the need to create adaptations of the test material, or extend tests with extra tasks prepared for students with a specific type of SEN (*alternate assessment*). This gives all students equal access to the content layer of the tasks and comparable opportunities to demonstrate their skills and knowledge, and teachers the opportunity to adequately assess the results by accurately establishing the relationship: individual – reference group (Ketterlin-Geller, 2005).

Universal design, both in the UDL and the UDA model, has become a principle consistent with the concept of educational inclusion (Spencer, 2011; Hall et al., 2012). Its scope covers, apart from teaching, other activities and services created for education and supporting the educational process. Psychological and pedagogical diagnosis is an inseparable element of education, especially in its inclusive form, in which every student is a subject, including the ones with special educational needs of various natures and intensities. Improving children's functioning in both the school and family environment entails recognizing their individual needs and abilities, and identifying and removing barriers in these environments. These activities are initiated by the diagnostic process and carried out by the school in cooperation with a psychological and pedagogical counseling center.

Universal design is part of the biopsychosocial model of disability, which is a guideline for educational activities in the field of psychological and pedagogical support, including comprehensive assessment. Nevertheless, in psychological practice, the application of UD principles to increase accessibility for all clients is considered relatively rarely: UD in the diagnosis and therapy for families (Bernal & Zera, 2012), UD in the assessment of children with reading difficulties (Braginets, 2018), and UD as a guideline for psychological services in APA Resolution on Support of Universal Design and Accessibility in Education, Training, and Practice (APA, 2019).

Therefore, it is necessary to postulate the use of the universal design model in the field by creating assessment tools and designing the process of psychological assessment. A set of guidelines for test developers as well as test users are the Standards for Educational and Psychological Testing issued by the American Educational Research Association [AERA]. A key change in the latest version of the standards so far "was the elevation of fairness as a foundational element of professional testing and assessment practice" (Jonson & Geinsinger, 2022, p. 4). The test can be made accessible through both adaptation (a reactive process) and universal design (a proactive process). Accessibility which is a legal requirement in some testing contexts means that "all test takers should have an unobstructed opportunity to demonstrate their standing on the construct(s) being measured" (AERA, 2014, p.49). Universal design is an approach to test development that seeks to maximize accessibility for all potential test takers. Developing a test according to the principles of universal design should be focused on maximizing fairness: "Universal design emphasizes the need to develop tests that are as usable as possible for all test takers in the intended test population, regardless of characteristics such as gender, age, language background, culture, socioeconomic status, or disability" (AERA, 2014, p.57).

This approach entails creating various forms of the diagnostic tool, giving the assessed person the opportunity to solve diagnostic tasks in an accessible way, and the diagnostician the chance to identify not only this person's needs but also their abilities. This is an important component of diagnostic practice, especially with respect to students with special educational needs to ensure the objectivity of measurement. Its increase is served by Universal Design and related accessibility. UD is designed to enable the examinee to respond to test tasks in a way that is accessible to them,

and for the diagnostician to ensure the best possible understanding of the examinee's needs and abilities.

In this context, a question arises: Is it possible to conduct a psychological assessment of students with different developmental needs using one universally designed tool (without developing alternative versions)?

PRINCIPLES OF UNIVERSAL DESIGN - LITERATURE REVIEW

The universal design commonly follows seven principles of equal access. Initially, they were applied to the physical elements of the environment (Connell et al., 1997). Currently, they are interpreted in various contexts, in which not only material but also non-material effects of the activity are created for a wide range of recipients. One of the perspectives in the field of education is the reference of the basic principles of universal design to assessment in education (Thompson et al., 2002), another perspective is required by the diagnostic process as an important stage of psychological and pedagogical counseling, considering both the tools and the diagnostic procedure. Examples of the application of UD principles in test construction are shown in Table 1.

Table 1. **Examples of the application of UD principles in test construction**

Principle	General description in the context of assessment	Application in test construction
Principle of Equitable Use	Designing tests and tasks in such a way that they are equally accessible to all students, including those with various special educational needs. Due to this, the differences in the results are not caused by variables not directly related to the measured construct.	At the stage of conceptualizing a diagnostic test, the broadest possible range of needs of various groups of students should be considered in terms of access to: – the form of test tasks (visual, auditory, and linguistic accessibility) e.g. tasks presented in different font sizes, graphics, volume or tempo of spoken text and audio, linguistically accessible instructions without culturally loaded terms; – the content of test/items (cognitive accessibility) e.g. multi-level tasks with varying difficulty levels; – procedures e.g. adjusted time limit or no time limit. Limitations: The need to control the impact of the adjustments on the accuracy of the measurement of a specific construct.
Principle of Flexibility in Use	The diagnostic tool should be usable in various forms allowing the examinee to select adaptations according to their individual preferences.	Examples of flexible adjustments allowed in various types of tests: – variable font size (Allan, 2003), – multimodal access to test content (Dolan et al., 2005), – extended response time, – replacing a motor response (indication) by a verbal response (Braden & Elliot, 2003), – a variety of language forms: a sign system, speech, simultaneous communication, fingerspelling (Easterbrooks et al., 2015), – translating instructions into a foreign language used by the diagnostician in performance tasks without cultural references (Barzykowski et al., 2013), – using with assistive technologies (e.g. answering questions only using the keyboard in digital tests. The principle of flexibility in use is easier to implement in computer tests due to the greater possibility of modifying the audiovisual layer of the test. Limitations: The limit of applying this principle in psychological diagnosis is the necessity to meet one of the formal criteria of the test: standardization – maintaining uniform test conditions.
Principle of Simple and Intuitive Use	All the elements of the test related to how it is used both by examinees and diagnosticians should be available, easily digestible and independent of their knowledge, experience, or skills.	Examples of principle: - task instructions and questions expressed in clear, precise language so that there is no need to use additional explanations, - ease of use of the test - which is related to the consistency of the procedures for starting and completing the test and the consistent use of the elements navigating the diagnostician and the examinee in the entire battery.

Principle of Simple and Intuitive Use	All the elements of the test related to how it is used both by examinees and diagnosticians should be available, easily digestible and independent of their knowledge, experience, or skills.	In computer-based tests, it is helpful to adhere to this principle by implementing the WCAG criteria, which support the intuitive use of a diagnostic tool, e.g.: - descriptive transcription displayed on the screen to all audio materials that are not test stimuli, - a logical and intuitive navigation sequence (consistency in the use of buttons to move forward and go back, exit, and terminate the test, - information about the success or failure of the training tasks communicated visually and audibly, - avoiding instructions referring to the shape, size, or location of the elements (e.g. click the square icon to continue), - unlimited time to read the instructions (W3C, 2008).
The Principle of Perceptible Information	Any information provided to the examinee should be fully comprehensible and easily noticeable.	This principle applies to messages such as instructions, feedback, descriptions of sample or training items, commands informing about the transition to the next section, and tasks. The effectiveness of the message should be enhanced in terms of perception for instance by using: - multimodal presentation in a graphical, textual, verbal, audible, and procedural form, - a division of information into smaller blocks, - test tasks preceded by several training and introductory tasks (CAST, 2008). When designing the graphics of a universal the following dimensions of the test's graphical readability should be considered: - font size and type (large print: 14 points, sans-serif of equal width), - contrast (white matte or pastel background), - spacing (1,25 points), - illustrations and drawings (avoiding grayscale and combining green and red colours, editing to ensure clarity but maintaining the so-called distractors and elements whose removal may make the task unjustifiably easier (Thomson et al., 2004).
The Principle of Tolerance for Errors	Such a construction of the tool that mistakes can be corrected or reverted by the examinee.	A universally designed test should contain training items in which the individual receives feedback on how they have performed on the training task. Since additional hints and feedback during the test tasks stage are not allowed, task instructions should be clearly formulated, without overburdening the immediate memory if other functions are tested. In computer-based tests, this principle requires to design and develop the tool in a way that minimizes the risk of making unintentional errors resulting from test malfunctions. Limitations: Applying the principle of tolerance to errors in psychological assessment is not always possible. In tasks where the reaction time is the indicator of the measured property, allowing the student to correct their mistakes interferes with the tested variable - the reaction speed.
The Principle of Low Physical Effort	This principle is especially important for students with special educational needs, including disabilities, who may experience higher levels of fatigue for various reasons: slower information processing, the use of specialized aids or assistive technologies, or disruptions in the perception of content in the modality affected by the deficit.	This principle can be implemented in many ways depending on the nature of the task and the purpose of the measurement, for instance: - ensuring breaks during the test according to the student's individual needs, - dividing the test procedure into shorter parts, - reducing the number of items, - increasing the time limit for performing tasks, - introducing rules for terminating the test.

The Principle of Size and Space for Approach and Use	The principle describes the need to ensure that all examinees can use the test material.	The test should be designed in such a way that giving answers, filling in form fields, and completing tasks are feasible for all students. Limited computer skills or the inability to physically operate a mouse or keyboard can be a barrier in computer-based testing. Therefore, the permission to replace motor responses with verbal ones to indicate a choice can be included as a solution acceptable in diagnosing psychological adaptation to the needs of people with SEN. The test design should also support situations when it will be taken by people using assistive technologies (e.g. the blind or people with physical disabilities).

Properly applied UD principles allow students with diverse learning needs to use diagnostic tools on equal terms with others. At the same time, the scope of possible modifications does not result in the risk of developing a completely new test that would call into question the diagnosis results.

Universal Design And Its Application In Psychological Assessment: Test Development And Conducting Assessment

Valid and fair psychological or educational assessment of students with special educational needs requires minimizing construct-irrelevant barriers for all examinees. Usually, test adaptations serve this purpose but thanks to universal design, we avoid the need for secondary modifications to the tool

Based on the Standards for Educational and Psychological Testing (AERA, 2014) and principles of test construction, we distinguished 4 four levels on which carefully thought-out actions are necessary to apply in psychological assessment: the conceptual, formal, test administration, and interpretive level.

The conceptual level of a universally designed test is based on a precise conceptualization of the properties measured by the test, Content-wise constructs should be justified in the meta-analyses of the literature on a given area, and the confirmed accuracy and predictive value of the results based on empirical evidence (*evidence-based practice,* EBP) (Paluchowski, 2010). The operationalization of the target construct makes a clear distinction between significant and non-significant sources of its variability. At this level, the author od a test should also decide whether certain ability is present in the target population at all, and if so, how it is specific to a given SEN category (Szubielska, 2017). For a given test to be used with students with different qualitative and quantitative cognitive characteristics, the *cognitive complexity* of a given task requires detailed analysis (Ketterlin-Geller, 2008). At the stage of conceptualizing the constructs measured by the test, its authors should assess whether the skills necessary to perform the task used in parallel with the *target skill* do not constitute a factor disturbing or limiting performance at the task. Secondly, a universally designed tool can be used to measure a wide range of specific properties, without excluding groups with special needs sanctioning their significantly lower or higher level (at least in theory). Thus, there may be a need to create multilevel scales with retraction rules that are justified by the psychometric analysis of position parameters. A wide range of task difficulty levels allows diagnosticians to obtain information also about the strengths of children who achieve the so-called "floor effect" in traditional tests (Hessl et al., 2009) making it possible to indicate the strengths of students with strongly differentiated capabilities in a tested property. The conceptual structure of a universally designed test also consists of items devoid of *culture bias*, *gender bias* or *disability bias* (Reynolds & Suzuki, 2013). In this aspect, it is important to maintain linguistic transparency both in the instructions and in the content of tasks that measure properties other than linguistic competence. Universal design takes into account not only the similarities of the needs and capabilities of all students but above all the differences that may modify the measurement of the assessed property. The possibilities of flexible adjustments implemented in the test structure (e.g. testing the same construct in a deficit-free modality) aim to eliminate this influence, giving all students an equal chance to examine specific properties (Dolan et al., 2005).

The formal level of the test includes the physical properties of universally designed tasks, graphics, and sound enabling correct reception of the content. The universal design of the test in the formal layer involves eliminating barriers to access the test: cognitive (e.g. too difficult tasks), sensory (e.g. blurred text or distorted sound), physical (e.g. inability to answer in the available modal-

ity), and linguistic (e.g., incomprehensible instructions) (cf. Thompson et al., 2004). Examples of solutions recommended at this level are adapted or modifiable font properties like color and contrast, perceptual availability of graphic and audio elements, or an alternative in the form of various forms of linguistic communication: text, sound, sign language, Braille language system, and translation into a foreign language as long as these solutions leave the measured property intact. Empirical research confirms that changing the physical properties of a task may facilitate the subject's perception of the content essential for performing the task correctly. For example, adaptable font properties when reading text on a computer screen turned out to be a better solution for visually impaired youth than presenting the text in enlarged print – the students could read faster, showed a similar level of understanding, and made fewer mistakes (McLaughlin & Kamei-Hannan, 2018). On the other hand, using a cream background alleviates the effects of a specific perceptual deficit, the so-called *scotopic sensitivity syndrome* in people with dyslexia and low vision (Kriss & Evans, 2005). If the computer or the diagnostician read the text of the task out loud, students with learning disabilities can understand the task instructions and its content, and poor reading skills will not affect the measurement of the assessed property when it is not reading literacy (Calhoon et al., 2000; Sireci et al., 2003; Dolan et al. 2005). The literature emphasizes that computer-based text-to-speech conversion is more effective than the diagnostician reading the text, especially if the text is long (more than 100 words) (Strangman & Dalton, 2005; Dolan et al., 2005).

Test administration level of the test in the universal design model assumes adjustments in the course of the test, such as extending time limits, allowing additional breaks, dividing the test into stages, and providing various ways of answering the questions. People with disabilities process information more slowly, use additional external aids, and have limited possibilities of accessing the task, adjusting their reaction, and presenting the tested skill. Extending the time limit is one of the adjustments reducing these barriers but due to the varied effects among people with the same type of disability, it should not be fixed but flexibly adjusted to individual needs (Lindstrom, 2010; Lovett, 2011). Extending the time limit for performing a task for people with disabilities cannot in any way favor this group, and the benefits of this adaptation should not be higher than just providing access to the task on conditions comparable to those for non-disabled people (Sokal & Vermette, 2017). In addition, to the extent allowed by the specificity of the task, the procedure should provide as wide a range of options for answering questions as possible: verbally, in writing, using the keyboard, by indicating a choice, considering the diverse ways of responding to specific stimuli by people with mainly sensory or mobility disabilities. According to the guidelines for implementing general standards in the field of diagnostics of people with disabilities (PPA, 2018), the psychologist is to ensure that the test is conducted only in standard conditions, but at the same time to provide people with disabilities access to diagnostic psychological tests through, among others, adjusting the test time, or using only parts of the test. The role of the psychologist is to assess the disruptive role of skill limitations and on this basis to determine whether the adjustment of the course of the test is appropriate in a specific situation and what impact it has. This document emphasizes the fact that people with the same type of disability may differ significantly in what adaptations they need, therefore this process should directly correspond to the specific needs of the diagnosed person.

The interpretive level is an important element that requires careful analysis during universal test design. It entails considering whether the referenced standardization group is representative: and if it is an appropriate point of reference in the case of a specific student. The problem in interpreting the results of an assessed person with a disability using a test standardized with the non-disabled population is unknown accuracy and reliability. It can therefore be postulated that the normalization sample should also include a balanced fraction of children representing specific groups with SEN. Introducing double norms: for the standardization group and the subgroup with SEN will allow diagnosticians to describe the functioning of the child concerning two reference groups: general and specific (interindividual comparisons). According to the Guidelines for implementing general standards in the field of diagnosing people with disabilities (PPA, 2018), it is the psychologist who assesses whether, from the perspective of answering the basic diagnostic question, it is more important for them to obtain information on how a given person performed against the general population, a specific clinical group, or a group of people with the same type of disability. Profile analysis is particularly useful in interpreting the results obtained by people with disabilities. Creating profiles of the distribution of results for individual groups, including the specificity of the problem and appropriate SEN subtypes, allows diagnosticians to properly orientate their diagnos-

tic conclusions. In addition, the developmental specificity of people with special educational needs as well as the intra-group heterogeneity of their functioning mean that verifying the significance of intergroup differences in terms of individual test results may be of less value than the assessment of certain properties against others in the same person (intraindividual comparisons).

THE APPLICATION OF UD RULES IN COMPREHENSIVE ANALYSIS OF COGNITIVE PROCESSES KAPP

The application of UD rules in the creation of psychological testing takes into account the mentioned 4 levels, This is illustrated using the example of a test battery called the Comprehensive Analysis of Cognitive Processes – KAPP (Comprehensive Analysis of Cognitive Processes – KAPP (acronym in Polish) (Bedyńska et al., 2021; Krasowicz-Kupis et al., 2022). The battery is used for functional diagnosis of cognitive functions including executive functions and language and communication. According to the theoretical assumptions, the battery was constructed with a view to its use in the diagnosis of people with special educational needs from groups of neurodevelopmental disorders and disabilities. Such an assumption was aimed at enabling psychologists to carry out a diagnosis of the cognitive sphere in people with different perceptual dysfunctions, various levels of intellectual performance and experiencing specific difficulties related to a given type of neurodevelopmental disorder or the situation in which the person finds himself (students for whom the language of school education is not the first language).

For this purpose, the principles of universal design and WCAG were used in constructing the test. These principles were implemented to the extent possible to preserve the theoretical relevance of the tool and the objectivity associated with maximally similar test conditions. When constructing the battery at the formal level, care was taken to ensure the accessibility of the tool in both the perceptual and content dimensions. As for the first type of accessibility, it was ensured through the use of sans-serif font in the instructions and items, a cream background, the possibility of changing the font size, reversing the contrast, and the use of illustrations. In tasks where this does not interfere with the measured construct, multi-modal access to the content of the test was used through dual (auditory and visual) instructions. Task instructions are expressed in clear, precise language - so that there is no need for additional explanations. The KAPP also provides special training in the use of the mouse and keyboard for subjects who need it (before the actual test begins). To enable proper understanding of test instructions, sign language was used in their presentation. In addition, the instructions were translated into three foreign languages: English, Ukrainian, and Vietnamese.

In the content layer (content accessibility), care was taken to vary the level of difficulty of the tasks, and the standardized testing procedure provides for students from certain groups to use tests corresponding to a younger age group. Such a situation is encountered, for example, in the procedure for assessing students with experience of migration, for whom Polish is the language of education, but not the mother tongue. Planning adequate educational activities in this case is possible after recognizing the level of specific skills of the child (e.g., reading in Polish). In such a situation, it is beneficial to use tests relating to different levels of development of a given skill, and not only the level consistent with age.

The level of test administration takes into account the limitations of the subjects, which usually in tests do not allow a proper assessment of the individual's level of ability. The most important of these is the restrictive time limit for completing tasks. In the KAPP battery, the time of presentation of the stimuli (for example, in the Reading with Comprehension test) and the time limits for answering have been set so that the factor, which in this case is reading speed, does not affect the assessment of reading comprehension. At this level, the possibility of choosing adjustments in the test procedure following the individual preferences of the examinee was also provided, such as the possibility of presenting instructions or items in unison (turning off the reader), the possibility of modifying the audiovisual features of the text: font and contrast, or the presentation of instructions in foreign languages. The scope and conditions under which the examinee can take advantage of the adjustments have been clearly defined so as to minimize their impact on the variability of the results, thus ensuring that the conditions for standardization are maintained.

DISCUSSION

Benefits of Using the UD Approach for Psychological Assessment

The universal design of psychological tests and the entire test procedure is a solution conducive to formulating a functional assessment for education. It brings more benefits for creating plans of support activities implemented at school than the information obtained from

a diagnosis based on traditional diagnostic instruments. The diagnostic conclusions obtained from a universally designed test, intended for a wide group of children, provide the basis for conducting a much more unified and at the same time individualized methodological/therapeutic process, using the similarity of educational needs classified into various categories of disability, which is so important in inclusive education. It becomes feasible to translate diagnostic descriptions into specific activities carried out at school. Universally planned diagnosis reduces the problem of diverse terms and specialized language codes, which facilitates intersubjective communication between the members of the team planning the interventions.

According to Messick (1995) what needs to be valid in educational testing is the meaning of the test scores as it entails further therapeutic actions. When the tool does not align with the capabilities of the individual being assessed, there is a high probability that the results will be distorted. A universally designed diagnostic tool eliminates at least two reasons for the loss of measurement accuracy of certain properties in children with disabilities. The first cause of disturbances in the accuracy of tests and procedures in assessing people with disabilities concerns the variance of test results, which is not related to differences in the level of the tested skills and properties but results from the cognitive, sensory, physical, and linguistic barriers in accessing the task. For example, unclear instructions make it impossible to solve the task correctly, and too small font causes additional perceptual effort that may reduce the examinee's efficiency when performing the actual task. The second possible reason for the loss of validity is the underrepresentation of the property measured by the test by excluding from the study sensually or cognitively unavailable scales (Braden & Elliot 2003; Spurgeon 2017).

An additional advantage of following the principles of UD in designing psychological assessment measures is the ease of using such instruments by the diagnostician. Apart from being aware of the disturbances of accurate measurement resulting directly from disability, the psychologist should know how to adjust the conditions and the course of the diagnostic test to the needs and capabilities of a person with a disability. Their role is also to identify the fine line between adaptation (deviation from the standard test procedure, which does not entail a change of the measured construct) and modification (deviation from the standard test procedure, which results in a change of the tested construct and thus incomparability of the results with the original tool). A universally designed tool allows test authors, at least partially, to ignore the controversy related to the choice of an adjustment and the assessment of its impact on the reliability of test results or the accuracy of conclusions drawn on their basis. A tool adjusted to make it available to the widest possible range of individuals becomes a solution friendly to the diagnostician and the student with significantly increased quality of the diagnosis. Thanks to this, diagnosticians can significantly limit the group of students excluded from the diagnostic process due to insufficient "diagnosability" of those with visual impairments (Atkins, 2011), hearing impairments (Maller, 2003), neurodevelopmental disorders (Tenorio et al. 2014), autism spectrum disorders (Courchesne et al., 2019), and those who are ethnically, linguistically, and culturally different (Barzykowski et al., 2013). Persons classified in a given SEN category have a heterogeneous spectrum of needs and abilities within their group, resulting in unequal access to the content of test tasks and different possibilities of adhering to standard test conditions. A rigid assumption about the need for a specific type of adaptation could be detrimental to some of them (e.g. that all visually impaired people need to zoom in), and algorithmic treatment of each student with a given type of SEN (e.g. dyslexia, intellectual disability) could limit test conclusions to information on symptomatic difficulties for a given SEN category. These adjustment problems and uncertainties also decrease when a test is universally designed.

If UD principles are used to create test scales in the content or interpretation layer (intraindividual variability, general and group-specific norms) diagnosticians will be able to obtain information also about the strengths of children with SEN, which is especially useful in designing interventions implemented at school, such as educational and therapeutic activities (Weishaar, 2010; Elder et al., 2018), constructing strategies in the field of psychological counseling (Magyar-Moe et al., 2015), and supporting activities aimed at stimulating students' strengths (Cosden et al., 2006).

A universally designed diagnosis enables flexible and dynamic examination, as it takes into account the variability of skills that are also assessed from the perspective of an individual (Lebeer et al., 2013). The impact of learning, development, and planned interventions, as well as the unstable nature of the child's disorders, needs, and abilities, mean changing needs in terms of adjusting the form of the test, procedure, and conditions for the same student (different levels of fatigue, changing needs in terms of the time needed to perform the task,

changing preferences in terms of modality when answering questions, changing needs related to material presentation such as size, contrast, volume, and audio speed). A universally designed tool is sensitive to such changes, and thus it enables psychologists to formulate conclusions about the developmental progress of the student in a longitudinal perspective.

Limitations of the UD approach in psychological assessment

The essence and the greatest difficulty of universal design is to create a tool flexible enough to maintain the integrity of the tested constructs and at the same time take into account the special perceptual and cognitive needs of a diverse population of examinees aiming at an unbiased assessment (Ketterlin-Geller, 2008; Szubielska, 2017). The aim of the test should be to maximize the individual's chances of demonstrating their ability and skill in the measured area without compromising the accuracy of the conclusions. This is connected with the necessity to precisely consider possible sources of diagnosis bias in terms of the theoretical construct, the method, and the test items (Hornowska & Paluchowski, 2004) at the stage of designing the tool in the UD model. In psychological assessment – a process with a significant and long-term impact on the school, family, and life situation of a child – this risk of biased interpretation should be mitigated. Psychological tests need to meet formal criteria, i.e. standardization (uniform test conditions), reliability (how accurately it measures the variable), and normalization (the possibility of referring the raw score to test standards informing about the typical test performance in a reference group). The key issue is the validity of the test, i.e. the certainty that it measures the right construct. Depending not only on the diversity of the needs of the test population but also on the nature of the measured properties, it is possible or not to meet the above-mentioned criteria in a universally designed test. In the case of tests assessing cognitive development, the presentation of the test material, its form, and content may be of key importance for the measured mental property, therefore it is not always possible to make changes to this material. This creates the risk that the test will lose its diagnostic usefulness since the measurement will not comply with the adopted theoretical assumptions. Empirical studies assessing the impact of using specific adjustments on the variability of results in a diagnostic test show fragmentary and heterogeneous results (Dial &Dial, 2010). This means that even in a universally designed test, with a wide spectrum of adjustments (such forms of deviation from the standard test procedure that do not entail changes to the measured construct), it may be necessary to introduce modifications of the tested construct that will make it impossible to compare results (PPA, 2018). Thus, the limitations of universal design in psychological diagnosis can be considered in two ways.

The first is the creation of such a repertoire of adjustments to the physical characteristics of tasks and items (formal level), which only reduces the impact of factors independent of the measured property (related to disability) and enables its accurate measurement and comparison of the results to the reference group and norms. Such adjustments are intended to allow equal access to the task for all subjects, but at the same time, they do not change the nature of the measured construct. Thus, a universally designed test cannot be method-biased. For example, it is not possible to translate graphic elements in reading comprehension tests (e.g. matching a caption to an image) into audio-descriptive text or tactile graphics to accommodate the needs and capabilities of blind people. A task modified in this way would not be equivalent to its visual version. In the first case, verbal information is related to the answer, whereas in the second case the tactile graphics and the diverse skills of its reception by the blind person. The version of the test in which test items are read and marked by a diagnostician may also be considered non-equivalent, if originally the subject performs these activities independently, due to the greater load on working memory and a stronger effect of the variable of social approval (Szubielska, 2017). Similar limitations in implementing the adaptations related to the WCAG accessibility criteria are noticed in computer-based tests. Not all accessibility rules may apply in psychological assessment. For example, providing instructions that screen readers can see are functions that cannot be used in a reading ability test since a situation when a screen reader reads a text presented as handwriting no longer measures target decoding ability, but other auditory perception functions. In such cases, it is impracticable to make adjustments that would result in creating an equivalent version of the task. Universal design should be treated more broadly here – not only as a process resulting in a uniform set of equivalent tests with built-in adjustment options, enabling the diagnosis of the same areas of functioning in all the subjects, but as a process-oriented toward flexible assessment of a specific repertoire of skills and properties if it is necessary – using also parallel non-equivalent tests (e.g. an alternative auditory test to test cognitive functions for the blind instead of excluding the task altogether). Thanks to

this, the problem of underrepresentation of the measured property is ignored, and the diagnosis of people with specific types of disability, for whom only the available scales have been used selectively, can take the form of a holistic image of their functioning. The results of such a study, although they cannot be referred to the general norms, provide valuable information about the functioning of the individual against the background of the population with a specific type of SEN and in an individual, longitudinal, and developmental approach.

The second difficulty in designing universal psychological tests lies in creating tasks taking into account their cognitive complexity (conceptual level) and assessing whether the specificity of the development of a person does not modify one of the skills needed to perform the task. Abilities assessed in psychometric tests may differ in terms of structure (e.g. working memory structure in blind and sighted people) and functioning (understanding of a text read out loud by hearing and deaf people). The point is therefore to avoid the bias of the theoretical construct: to include in the measurement a property that does not characterize or characterize to a limited extent the population for which the diagnostic tool is adapted (Szubielska, 2017). An example is a task that tests the ability to switch attention, whose execution requires efficient analysis and visual synthesis, quick reactions (perception speed), and working memory, and with such complexity it disadvantages people with visual impairments or intellectual disabilities. Providing additional exposure time to the stimuli modifies the measurement, and makes the results less useful for drawing diagnostic conclusions. Another example of such a situation encountered in computer-based research may be deficiencies in computer knowledge and skills), which make it difficult to perform the test, or their low quality which interferes with measuring the target property (ITC, 2005). In order to minimize the impact of such limitations, appropriate solutions should be introduced, e.g. training in the use of the computer functions required in the test before the test, an intuitive interface, and a simplified way of coding answers by the subject.

There is also controversy around the practice of flagging the results obtained with a specific adjustment (cf. Ketterlin-Geller, 2008). On the one hand, the purpose of adjustments is to eliminate only distortions resulting from the disability and to maintain a uniform target construct. On the other hand, in psychological research, any change in standard conditions may modify diagnostic conclusions, which are the basis for formulating long-term directions of support, and therapeutic and educational activities. Both the minimal change in the test conditions (the use of adjustments) and the lack of it in the case of people with disabilities may lead to a biased diagnosis (Szubielska, 2017).

CONCLUSION

- Universally designed tools create standard test conditions for people with various limitations, giving psychologists a chance to obtain accurate, reliable, and comparable test results for the vast majority of respondents. This makes it possible to carry out a comparative analysis between different groups of students with SEN and a diagnosis that does not reduce the measurement to selected aspects - possible due to the limitations in the functioning of a specific student.
- The flexibility assumed in UD poses a risk of reducing the goodness of the diagnosis (reliability and validity – the psychometric aspects of a tool).
- Our work on the development of the battery of cognitive tests described above (KAPP) allows us to conclude that the application of UD principles in the development of the test significantly broadens the range of potential individuals who can be assessed using it. The precise definition of the theoretical construct that is being measured, basing decisions on the rules of performing test tasks (e.g. time limitation) on the results of empirical studies ensures the accuracy of the measurement. However, the diversity of subjects with SEN and disabilities requires a flexible approach to diagnosis and, consequently, adaptation of tests and adjustments to the individual characteristics of the recipients.
- Enabling individuals with disabilities to undergo accurate and reliable diagnosis requires the application of a UD (Universal Design) approach at four levels: conceptual, formal, test administration, and interpretative. Selective application of UD principles will not eliminate the impact of barriers on result variability.

ACKNOWLEDGEMENT

None

DECLARATION OF INTEREST STATEMENT

The author reported no potential conflict of interest.

FUNDING

None

REFERENCES

Allan, J. (2003). Editorial social inclusion. *Policy Futures in Education*, *1*(4), 622-625.

AERA, (2014). *Standards for educational and psychological testing*. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

APA, (2019). *Resolution on Support of Universal Design and Accessibility in Education, Training, and Practice*. American Psychological Association. Retrieved from https://www.apa.org/about/policy/universal-design-accessibility-education (accessed September 21, 2023).

Atkins, S. (2011). *Assessing the ability of blind and partially sighted people: Are psychometric tests fair?* Birmingham: RNIB Centre for Accessible Information.

Barzykowski, K., Grzymała-Moszczyńska, H., Dzida, D., Grzymała-Moszczyńska, J., & Kosno, M. (2013). *Wybrane zagadnienia diagnozy psychologicznej dzieci i młodzieży w kontekście wielokulturowości oraz wielojęzyczności* [*Selected issues of psychological assessment of children and adolescents in the context of multiculturalism and multilingualism*]. Warsaw: Ośrodek Rozwoju Edukacji.

Bedyńska, S., Campfield, D., Kaczan, R., Kaczmarek, M., Knopik, T., Kochańska, M., Krasowicz-Kupis, G., Krejtz, I., Orylska, A., Papuda-Dolińska, B., Rycielski, P., Rydzewska, K., Sędek, G., Smoczyńska, M., & Wiejak, K. (2021). Narzędzia diagnostyczne do oceny zdolności poznawczych dzieci i młodzieży – projekt wdrożeniowy [Diagnostic tools for assessment of cognitive functioning in children and youth–the implementation project]. *Przegląd Psychologiczny*, *64*(2), 9–23. https://doi.org/10.31648/pp.7323

Bernal, A. T., & Zera, D. A. (2012). Increasing accessibility in couple and family therapy training: Incorporating universal design for instruction. *Contemporary Family Therapy: An International Journal, 34*(1), 112–123. https://doi.org/10.1007/s10591-012-9178-6

Bowe, F.G. (2000). *Universal Design in Education: Teaching nontraditional students.* Westport, CT: Bergin &Garvey.

Braden, J. P., & Elliott S. N. (2003). Accommodations on the Stanford-Binet intelligence scales, fifth edition. In: G. Roid, (ed.) *Stanford-Binet Intelligence Scales, fifth edition, interpretive manual: Expanded guide to the interpretation of SB5 test results* (pp. 135–143). Itasca, IL: Riverside Publishing.

Braginets, E.I. (2018). The application of universal design evaluation for fair testing children with reading disorders. *Journal of Modern Foreign Psychology*, *vol. 7*(3), 28-37.

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of Computer-Based Test Accommodations on Mathematics Performance Assessments for Secondary Students with Learning Disabilities. *Learning Disability Quarterly*, *23*(4), 271–282. https://doi.org/10.2307/1511349

CAST, (2008). *Universal design for learning guidelines 1.0*. Wakefield, MA: Author.

Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., Sanford, J., Steinfeld, E., Story, M., & Vanderheiden, G. (1997). *The principles of universal design*: *Version 2.0.* The Center for Universal Design. Retrieved from https://www.sphassociates.ca/uploads/files/Principles%20of%20Universal%20Design.pdf (accessed October 27, 2022).

Cosden, M., Koegel, L., K., Koegel, R. L., Greenwell, A., & Klein, E. (2006). Strength-based assessment for children with autism spectrum disorders. *Research & Practice for Persons with Severe Disabilities, 31*(2),134–143. https://doi.org/10.1177/154079690603100206

Courchesne, E., Pramparo, T., Gazestani, V. H., Lombardo, M.V., Pierce, K., & Lewis, N. E. (2019). The ASD Living Biology: from cell proliferation to clinical phenotype. *Molecular Psychiatry*, *24*(1), 88-107. https://doi.org/10.1038/s41380-018-0056-y

Dial, J. G., & Dial, C. L. (2010). Assessing and intervening with visually impaired children and adolescents. In D. C. Miller (ed.) *Best Practices in School Neuropsychology: Guidelines for Effective Practice, Assessment, and Evidence-Based Intervention* (pp. 465-479). Hoboken, NY: John Wiley & Sons, Inc.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment, 3*(7), 4-32.

Easterbrooks, M. A., Katz, R. C., Kotake, C., Stelmach, N. P., & Chaudhuri, J. H. (2015). Intimate partner violence in the first two years of life: Implications for toddlers' behavior regulation. *Journal of Interpersonal Violence*, *33*(7), 1192-1214. https://doi.org/10.1177/0886260515614562

Elder, B.C., Rood, C.E., & Damiani, M.L. (2018). Writing strength-based IEPs for students with disabilities in inclusive classrooms. *International Journal of Whole Schooling, 14*(1), 116-153.

Hall, T. E., Meyer, A., & Rose, D. H. (2012). *Universal design for learning in the classroom: Practical Applications*. New York: The Guilford Press.

Hessl, D., Nguyen, D. V., Green, C., Chavez, A., Tassone, F., Hagerman, R. J., Senturk, D., Schneider, A., Lightbody, A., Reiss, A. L., & Hall, S. (2009). A solution to limitations of cognitive testing in children with intellectual disabilities: the case of fragile X syndrome. *Journal of Neurodevelopmental Disorders*, *1*(1), 33-45. https://doi.org/10.1007/s11689-008-9001-8

Hornowska, E., & Paluchowski, W. J. (2004). Kulturowa adaptacja testów psychologicznych [Cultural adaptation of psychological tests]. In J. Brzeziński (ed.), *Metodologia badań psychologicznych. Wybór tekstów [A Psychological research Metodology. A Selection of texts]*, (pp. 151-191). Warsaw: PWN.

ITC, (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. International Test Commission. Retrieved from https://www.intestcom.org/files/guideline_computer_based_testing.pdf (accessed September 23, 2021).

Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). University of Minnesota, National Center on Educational Outcomes. Retrieved from: http://cehd.umn.edu/ (accessed August 12, 2023).

Jonson, L., & Geinsinger, K. F. (2022). Introduction. In L. Jonson, & K.F. Geinsinger, (eds.), *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standard,* (pp. 1-10). American Educational Research Association. https://doi.org/10.2307/j.ctv2kzv0fw

Ketterlin-Geller, L. R. (2005). Knowing What All Students Know: Procedures for Developing Universal Design for Assessment. *The Journal of Technology, Learning and Assessment*, *4*(2), 4-22.

Ketterlin-Geller, L. R. (2008). Testing Students with Special Needs: A Model for Understanding the Interaction Between Assessment and Student Characteristics in a Universally Designed Environment. *Educational Measurement: Issues and Practice, 27*, 3-16. https://doi.org/10.1111/j.1745-3992.2008.00124.x

Krasowicz-Kupis, G., Wiejak, K., Bedyńska, S., Błaszczak, A., Kaczan, R., Kaczmarek, M., Kochańska, M., Knopik, T., Krejtz, I., Łuniewska- Etenkowska, M., Orylska, A., Papuda-Dolińska, B., Rycielski, P., Rydzewska, K., Sędek, G., Smoczyńska, M., & Zieliński, P. (2022). *Kompleksowa Analiza Procesów Poznawczych – bateria diagnostyczna KAPP dla dzieci i młodzieży ze specjalnymi potrzebami edukacyjnymi [Comprehensive Analysis of Cognitive Processes – assessment battery for students with special educational needs]*. Warsaw: Instytut Badań Edukacyjnych.

Kriss, I., & Evans, B. J. W. (2005). The Relationship between Dyslexia and Meares-Irlen Syndrome. *Journal of Research in Reading*, *28*, 350–364. https://doi.org/10.1111/j.1467-9817.2005.00274.x

Lebeer, J., Partanen, P., Candeias, A., Grácio, M. L., Bohacs, K., Sønnesyn, G., Van de Veire, H ., Van Trimpont, I., Orban, R., János, R., Demeter, K ., Schraepen, B., & Dawson, L . (2013). The need for a more dynamic and ecological assessment of children experiencing barriers to learning to move towards inclusive education: A summary of results of the DAFFODIL project. *Erdélyi Pszichológiai Szemle, Spec Issue*, 175–205.

Lindstrom, J. H. (2010). Mathematics assessment accommodations: Implications of differential boost for students with learning disabilities. *Intervention in School and Clinic, 46*(1), 5–12. https://doi.org/10.1177/1053451210369517

Lovett, B. J. (2010). Extended Time Testing Accommodations for Students With Disabilities: Answers to Five Fundamental Questions. *Review of Educational Research*, *80*(4), 611-638. https://doi.org/10.3102/0034654310364063

Lovett, B. J. (2011). Extended time testing accommodations: What does the research say? *National Association of School Psychologists, 39*(8), 14–15. Retrieved from https://eric.ed.gov/?id=EJ934062 (accessed August 15, 2023).

Mace, R. (1985). Universal design: Barrier free environments for everyone. *Designers West, 33*(1), 147-152.

Magyar-Moe, J. L., Owens, R. L., & Conoley, C. W. (2015). Positive psychological interventions in counseling: What every counseling psychologist should know. *The Counseling Psychologist*, *43*(4), 508–557. https://doi.org/10.1177/0011000015573776

Maller, S. J. (2003). Intellectual assessment of deaf people: A critical review of core concepts and issues. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 451–463). Oxford. Oxford University Press.

McLaughlin, R., & Kamei-Hannan, C. (2018). Paper or Digital Text: Which Reading Medium is Best for Students with Visual Impairments? *Journal of Visual Impairment & Blindness*, *112*(4), 337-350. https://doi.org/10.1177/0145482X1811200401

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Nave, L. (2021). Universal Design for Learning UDL in online environments: The HOW of learning. *Journal of Developmental Education, 44*(3), 34–35.

Paluchowski, J. (2010). Diagnoza oparta na dowodach empirycznych – czy potrzebny jest „Polski Buros"? [Evidence-based assessment – do we need "polish Buros"?]. *Roczniki Psychologiczne, 13*(2), 7-27

PPA, (2018). *Guidelines for implementing general standards in the field of diagnosing people with disabilities*. Polish Psychological Association, Section of Psychological Diagnosis.

Reynolds, C. R., & Suzuki, L. A. (2013). Bias in psychological assessment: An empirical review and recommendations. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: assessment psychology,* (pp. 82–113). Hoboken, NJ: John Wiley & Sons, Inc.

Rose D. H., & Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria: Association for Supervision and Curriculum Development.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Center for Educational Assessment Research Report no. 485). Amherst, MA: School of Education, University of Massachusetts Amherst.

Sokal, L., & Vermette, L. A. (2017). Double time? Examining extended testing time accommodations (ETTA) in postsecondary settings. *Journal of Postsecondary Education and Disability*, *30*(2), 185–200.

Spencer, S. (2011). Universal Design for Learning: Assistance for Teachers in Today's Inclusive Classrooms. *Interdisciplinary Journal of Teaching and Learning*, *1*, 10-22.

Spurgeon, S.L. (2017). Evaluating the Unintended Consequences of Assessment Practices: Construct Irrelevance and Construct Underrepresentation, *Measurement and Evaluation in Counseling and Development, 50*(4), 275-281. https://doi.org/10.1080/07481756.2017.1339563

Strangman, N., & Dalton, B. (2005). Using technology to support struggling readers. In D. Edyburn, K. Higgins & R. Boone (Eds.), *Handbook of Special Education Technology Research and Practice*. Whitefish Bay, WI: Knowledge by Design, Inc.

Szubielska, M. (2017). Możliwości dostosowania testów psychometrycznych do specjalnych potrzeb rozwojowych i edukacyjnych badanych osób [Possibilities of adapting psychometric tests to the special developmental and educational needs of test subjects]. In K. Krakowiak (ed.) *Diagnoza specjalnych potrzeb rozwojowych i edukacyjnych dzieci i młodzieży. Standardy, wytyczne oraz wskazówki do przygotowania i adaptacji narzędzi diagnostycznych dla dzieci i młodzieży z wybranymi specjalnymi potrzebami rozwojowymi i edukacyjnymi [Assessment of special developmental and educational needs of children and adolescents. Standards, directions and guidelines for the preparation and adaptation of diagnostic tools for children and adolescents with selected special developmental and educational needs]* (pp. 104-118). Warsaw: Ośrodek Rozwoju Edukacji.

Tenorio, M., Campos, R., & Karmiloff-Smith, A. (2014). What standardized tests ignore when assessing individuals with neurodevelopmental disorders / *Lo que ignoran los tests estandarizados en la evaluación de personas con trastornos del neurodesarrollo*, Studies in Psychology, *35*(2), 426-437. https://doi.org/10.1080/02109395.2014.922264

Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments (Synthesis Report 4)*. University of Minnesota, National Center on Educational Outcomes.

Thompson, S., Thurlow, M., & Malouf, D.B. (2004). Creating Better Tests for Everyone through Universally Designed Assessments. *Journal of Applied Testing Technology, 6*, 1–15.

W3C, (2008). *Web Content Accessibility Guidelines (WCAG) 2.0*. Retrieved from https://www.w3.org/TR/2008/REC-WCAG20-20081211 (accessed September 21, 2023).

Weishaar, P. M. (2010). Twelve ways to incorporate strengths-based planning into the IEP process. *Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(6), 207-210. https://doi.org/10.1080/00098650903505381