Assessing intelligence in children and youth living in the Netherlands

Petra P. M. Hurks and Helen Bakker

Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands; Department of Developmental Psychology, Utrecht University, Utrecht, The Netherlands

ABSTRACT
In this article, we briefly describe the history of intelligence test use with children and youth in the Netherlands, explain which models of intelligence guide decisions about test use, and detail how intelligence tests are currently being used in Dutch school settings. Empirically supported and theoretical models studying the structure of human cognitive abilities, such as the Cattell-Horn-Carroll theory of intelligence (CHC) model, the model of primary mental abilities, and the Planning, Attention-Arousal, Simultaneous and Successive (PASS) theory of intelligence, are discussed in this context. Also, we explain who is allowed to administer intelligence tests in the schools in the Netherlands and what training and credentials are required to do so. Finally, some future directions for intelligence test development and use are discussed.

In the Netherlands, intelligence tests are included in many assessments of children and adults each year, for screening purposes, for determining educational or intervention needs and goals, and for evaluating treatment effects (Hurks, Hendriksen, Dek, & Kooij, 2013). The term intelligence can be defined as “the (cognitive) capacity of an individual to understand the world around him and his resourcefulness to cope with its challenges” (Wechsler, 1975, p. 139). In line with this, administering intelligence tests can be seen as a systematic study of this “capacity” using selected items (questions or assignments; Drenth & Sijtsma, 2006). Today, there are more than 20 individual-based and group-based tests for measuring intelligence available in Dutch. The majority of these tests can be used to measure intelligence in children and youth (for an overview, see Resing, 2015). The majority of these tests measure primarily the cognitive capacities of individuals, even though the authors of these tests often acknowledge the fact that noncognitive factors, such as motivation and perseverance, are necessary for being able to cope with the above-mentioned challenges in daily life as well. Some of the Dutch intelligence tests were originally developed in another language (primarily in English) and were translated, adapted, and normed for use in this country. Several other intelligence tests were developed in the Netherlands. In this paper, we will briefly describe the history of intelligence test use with children and youth in the Netherlands, the models of intelligence that guide decisions about test use and their interpretation, and how intelligence tests are currently used in the Dutch school settings.

History of intelligence test use with children and youth in the Netherlands

1900–1920s
In 1901, compulsory education for children and youth was introduced in the Netherlands. As a percentage of children could not keep up the pace of the (regular) educational programs offered, scientists and educators became increasingly focused on identifying those children who were in need of additional help or instruction with respect to the daily routine of primary education, and on determining what kind of help these children needed. One of these researchers was Dirk Herderschée (1877–1969). At the time, it was a thorn in his flesh that there were no psychometrically sound methods available for identifying those children who were in need of additional help or instruction with respect to the daily routine of primary education, and on determining what kind of help these children needed. One of these researchers was Dirk Herderschée (1877–1969). At the time, it was a thorn in his flesh that there were no
choice which children would be selected for extra help or special education programs. In this context, Herderschée translated and adapted, for one, the original version of the Binet-Simon intelligence test in Dutch (Baldi, 2014; Mulder & Heyting, 1998). Binet and Simon had published in 1905 a test battery measuring intelligence. They defined intelligence as “the ability to judge well, to understand well, to reason well” (Resing, 2015, p. 24). Individuals had to perform selected tasks that measured several cognitive aspects of intelligence, such as abstract reasoning, verbal comprehension, and working memory. Based on these tasks, a composite score could be calculated. Those children who scored below a specific cutoff point on this composite test score were, according to Binet and Simon, eligible for special needs education programs. The Dutch version of the Binet-Simon test (i.e., the Binet-Herderschée test) was published in 1919 and it was used, practically unaltered, until 1969 (Mulder & Heyting, 1998). This version of the test could be administered to children aged 5 years and older.

In 1911, the Amsterdam Paedological Society was founded, which was a Dutch society promoting studies in the field of pediatric medicine, child psychiatry, developmental psychology, and pedagogy. Members (such as Herderschée) were psychiatrists, physicians, psychologists, and pedagogues. This society played an important role in the introduction of intelligence testing in the Netherlands (Knegtmans & Kox, 2000). For one, at a conference organized by this society in 1913, early attempts to measure intelligence were discussed. At that point, the first studies investigating intelligence and specific cognitive functions, such as attention, in Dutch children had been conducted and were presented at the conference (Mulder & Heyting, 1998). For instance, Wiersma reported at the conference data revealing positive correlations among estimates of attention, intelligence, and school success in children. He also found a larger variability in performances on tests measuring attention and intelligence among boys than among girls (Mulder & Heyting, 1998).

During this conference, a lot of skepticism regarding the value of intelligence tests was expressed by members of the society. For instance, Schuyten claimed that the Binet-Simon intelligence test measured “knowledge” rather than “intelligence” (Depaepe, 1989, in Mulder & Heyting, 1998). Parallel to the conference, the Amsterdam Paedological Society published in 1913 two articles on the concept of “general intelligence” (or “g”), explaining (and discussing the validity of) the statements made by Charles Spearman, who introduced this concept in 1904. He had found that “all cognitive tests are positively correlated with one another, irrespective of the cognitive domain sampled” (Colom, Jung, & Haier, 2006, p. 1359). Based on this, Spearman concluded that g is “a general competence that is the foundation of specific cognitive abilities” (Mulder & Heyting, 1998, p. 355).

1920–1950

Wechsler published in the 1930s and 1940s two scales to measure intelligence: the Wechsler-Bellevue Intelligence Scale Form I (Wechsler, 1939) and Form II (Wechsler, 1946). Both scales formed the basis of the Wechsler tests that are frequently used worldwide today, including the Netherlands. Wechsler defined intelligence as “the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment” (Wechsler, 1944, p. 3). The aims of these tests were (and still are): (a) to estimate the “general intelligence” (or g) of children, youth, and adults, (b) to profile an individual’s cognitive strengths and weaknesses, and (c) to investigate the effectiveness of educational interventions. Based on the administration of these scales, several outcome measures could be calculated: a composite score (i.e., a total, or also called Full Scale, IQ score), as a measure of general intelligence, and several index scores to quantify specific cognitive functions (such as verbal intelligence, performance intelligence, and, later, processing speed). Unlike later versions of the Binet-Simon test discussed above, Wechsler constructed separate tests for (young) children and adults. These separate versions of the Wechsler test were (and still are) also available in Dutch, normed on a Dutch population. Note, however, that the most recently published English versions of the Wechsler batteries are not yet available in Dutch. Currently, only the third edition of the Wechsler test for primary school and preschool children (i.e., the WPPSI-III) and the third edition of the Wechsler test for older children and youth (the WISC-III) are available in Dutch, although new(er) versions of the Wechsler instruments (e.g., the fifth edition of the WISC and the fourth version of the WPPSI) are available, for example, in the United States. Due to the process of translating and norming, the Netherlands are always a few years behind when it comes to availability of the most recent versions of tests such as the Wechsler tests.

1960s–2010s

Parallel to the publications of the work of European scientists, for example, Rasch (1960) and Fischer (1974), the use and development of objective quantitative approaches, that is, psychological tests, was placed higher on the Dutch “education agenda” (Drenth & Sijtsma, 2006). To improve communication in this field, the Dutch Association of Psychologists founded, in 1959, the Commissie Testaangelegenheden Nederland (COTAN)
Evaluation system for test quality (in the past, this committee was called Test Research Committee; Evers, 2012). The main task of this committee was (and still is) to publish a summary of available Dutch psychological tests and to assess the quality of these tests. Since 1969, the COTAN has evaluated psychological tests, including intelligence tests, based on seven criteria including theoretical foundation(s) of the test, the quality of the test materials and the manual, norms, reliability (i.e., the extent to which measurements are repeatable), construct validity (i.e., does the test measure what it claims to measure), and criterion validity (i.e., whether a test score is a predictor of another, related behavior). By 2012, the COTAN had assessed and discussed more than 900 psychological tests and questionnaires (Evers, 2012). The number of tests that has been reviewed has expanded over the last years (Egberink, Janssen, & Vermeulen, 2009–2014). The majority of the 20+ Dutch intelligence tests discussed above have been reviewed by the COTAN. In the next paragraphs, we will discuss the models of intelligence that currently guide decisions about test use and their interpretation in the Netherlands and the intelligence tests most commonly used in this country.

Models of intelligence and decision making in the Netherlands

From the 1970s, test publishers increasingly started to construct tests in Dutch that (most often) measured not only one general intelligence (g), but also separate types or indices of intelligence, such as measures of working memory, fluid intelligence, crystallized intelligence, information processing speed, and so forth (Bos & Magez, 2016; Resing, 2015). Most of these intelligence tests also had one or more subtests that measured learning progress. Empirically supported and theoretical models studying the structure of human cognitive abilities (such as the CHC model, the model of primary mental abilities, and the PASS model) were (and are) used to develop these tests in the Netherlands. The tests, or combinations of tests, chosen based on these models allow practitioners to describe and to classify the individual’s observed characteristics (e.g., what are the specific strengths and weaknesses of the child or youngster) or to make predictions about the future (e.g., which school would be most suitable for the child or which child would benefit the most from special education).

The model of primary mental abilities

This model was postulated by Thurstone (1938) and states that a combination of specific cognitive (or primary mental) abilities determines “how well an individual understands the world around him and how resourceful he is in coping with its challenges.” More specifically, Thurstone defined nine primary mental abilities: verbal comprehension, verbal fluency, number facility, spatial visualization, memory, inductive reasoning, deductive reasoning, practical problem reasoning, and perceptual speed. Thurstone did not mention or acknowledge the concept of overall intelligence, or g, in his model. The nine primary mental abilities are believed to be equally important in determining or understanding the intelligence of an individual (i.e., these are not arranged in a certain manner; Matthews, Zeidner, & Roberts, 2004). This model also implies that each cognitive ability has to be tested in order to determine one’s intelligence.

The Cattell-Horn-Carroll theory, or CHC theory

Since Thurstone’s initial formulation of the model of primary mental abilities, many articles have been published discussing the number of specific cognitive functions that can and must be distinguished (and tested) to determine one’s intelligence. For instance, Carroll (1993) suggested that approximately 65–69 unique cognitive functions could be distinguished. However, it would be impossible, or at least extremely difficult, to test and take into account all of these abilities, while estimating one’s intelligence. Therefore, theoreticians have postulated alternative models defining these first-order specific (or primary) cognitive functions as well as so-called second-order factors, which are more broadly defined. In this context, the specific cognitive functions (or primary mental abilities) should be seen as elements of these second-order factors (e.g., Cattell, 1941, 1971). The Cattell-Horn-Carroll theory, or CHC theory, is one of these alternative models. The CHC theory defines factors on three levels (or strata): stratum I, which is called “narrow” abilities (i.e., the 65–69 specific cognitive functions mentioned above); stratum II, the second-order factors (in the CHC model, 10 second-order or “broad” factors are defined including crystallized intelligence, fluid intelligence, and quantitative reasoning); and stratum III, that is, one single factor measuring “general intellectual ability” (or g; Kaufman, Kaufman, & Plucker, 2013).

The Planning, Attention-Arousal, Simultaneous, and Successive (PASS) theory of intelligence

The Planning, Attention-Arousal, Simultaneous, and Successive (PASS) theory of intelligence (Naglieri & Das, 1997) is based on Luria’s (1966) neuropsychological model and defines three functional units (or blocks). The first block is planning: this unit includes higher-order, supervisory control functions that are involved in...
organizing behavior, strategy selection, and monitoring. The second block includes focused and sustained attention. The third block includes information processing abilities, in a simultaneous or sequential manner. Simultaneous processing is defined as the integration of information that is offered together at one time, for example, perceiving all elements of a photo at once. Sequential processing is the interpretation of information in a timely and successive order, such as remembering a telephone number (Kaufman, Kaufman, & Plucker, 2013). The PASS model states that all of these functional units need to be tested in order to estimate the intelligence of a person.

Finally, a cross-battery assessment approach is often used (Bos & Magne, in press; Flanagan, Ortiz, & Alfonso, 2013). This approach integrates neuropsychological theories (such as the PASS model) and intelligence theories (such as the CHC theory) and “provides practitioners with the means to make systematic, reliable, and theory-based interpretations of any ability battery and to augment that battery with cognitive, achievement, and neuropsychological subtests from other batteries to gain a more psychometrically defensible and complete understanding of an individual’s pattern of strengths and weaknesses” (Flanagan et al., 2013, p. 1). This allows a larger number of tests, which measure one or more cognitive functions, to be grouped to determine the intelligence level of an individual.

Development and use of intelligence tests in the Netherlands

As mentioned above, there are more than 20 tests available in the Netherlands for measuring intelligence. The majority of these tests can be used to measure intelligence in children and youth. Most of these Dutch tests are evaluated by the COTAN as having sufficient or good psychometric properties (Egberink, Janssen, & Vermeulen, 2009–2014; Resing, 2015). However, the majority of Dutch intelligence tests are less suitable for assessing children with limited fine motor abilities, visual impairments, or less well-developed language skills, for example, because the child has been living in the Netherlands for only a limited amount of time and obtained limited practice in communicating in Dutch.

International published tests (e.g., the Kaufman batteries, such as the Kaufman Intelligence Test for Adolescents and Adults, or KAIT; Kaufman & Kaufman, 2004; and the Wechsler batteries) have been translated and adapted for use in the Netherlands and are frequently administered. Also, several tests have been developed in the Netherlands for measuring intelligence in children and youth. Among those most widely used are the Snijders-Oomen Non-Verbal Intelligence Test (SON) series, that is, the SON-Revised (SON-R) 2.5–7 (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998; for children aged 2.5–7 years) and the SON-R 6–40 (Tellegen & Laros, 2011; for individuals aged 6–40 years). These tests include only performance subtasks and nonverbal reasoning subtasks and can be administered without the child or youngster having to use written or spoken language to respond. These tests do not include subtests measuring, for example, language (expressive or receptive) or verbal reasoning and are therefore often used with children and youth with language disorders, immigrant youth, and deaf or hearing-impaired children. Another example of a Dutch intelligence test that is frequently used is the Revised Amsterdam Intelligence Test (RAKIT-2; Resing, Bleichrodt, Drenth, & Zaal, 2012; for children aged 4–12.5 years), based on which a general intelligence score can be calculated as well as index scores for perceptual reasoning, verbal learning, nonverbal orientation, and verbal information processing speed. This test is less verbally loaded than the Weschsler scales and specifically suited for testing children with lower cognitive or intellectual capacities. Some tests used in the Netherlands explicitly aim to detect discrepancies between cognitive functions (or intelligence) and academic skills (e.g., the NSCCT test, or in Dutch the “Niet schoolse cognitieve capaciteitein test,” Van Batenburg & Van der Werf, 2004).

The intelligence tests seem relatively similar, partly because they claim to measure the same function(s), the theoretical foundations are highly comparable, and the choice of subtests is tuned in to the theoretical foundation, partly because they are quite traditionally constructed (with a history back to the 1950s), are an operationalization of models such as the CHC model (see above), or are being ad hoc (afterwards) interpreted in terms of these models. For instance, the third versions of the Wechsler batteries used in the Netherlands are increasingly being interpreted in terms of the CHC model even though these versions of the Wechsler tests were not constructed according to this model (Kaldenbach, 2006, 2007; Resing, 2015). However, subtle differences in the focus of the batteries can be observed as well, for example, to the extent to which verbal, nonverbal, fluid, or crystallized intelligence (or a combination of these) are being measured. Indeed, most intelligence tests contain multiple subtests, based on which subtest or indices scores can be calculated (in addition to an overall, general intelligence score) that provide information on the cognitive strengths and weaknesses of an individual and inform a clinician on whether an intervention or special education is needed and what the focus of this intervention or education program should be. The majority of these indices and subtests have reliable to good psychometric
properties. In practice, there sometimes is a discrepancy between the aim of the test (e.g., measuring general intelligence or general intelligence and some cognitive factors) and the tendency of psychologists and educators to want to make statements about strengths and weaknesses at a more specific level; for example, at the level of the subtest (Kaldenbach, 2006).

The majority of tests are paper-and-pencil tasks that are administered and scored individually, although, increasingly, test publishers are providing digital scoring tools and digital reporting tools. A few tests are available in Dutch that can be administered in a group situation. These tests are either administered as paper-and-pencil tests or on a computer. Examples are the Dutch ADIT test (in Dutch: Adaptieve Digitale Intelligentietest, Vegt & Metselaar, 2011) and the Dutch IVO test (in Dutch: Instaptoets VO, Verweij, 2002, 2008). These tests can be administered in groups consisting of 20 to 25 individuals. Group-based administrations have an advantage over individual-based testing: they are efficient and less expensive and the total testing time is in general less than that of an individually administered test. The disadvantage of a group-based administration is that there is less room for observing the individual during the test administration (Resing, 2015). Group tests are especially used for decision making regarding secondary school placement.

The use of intelligence testing in school settings in the Netherlands

In the Netherlands, compulsory education starts at age 5 years. At age 4, children can get enrolled, and the majority of children do. Between the ages of 16 and 18 years, education is only partially compulsory, which means that youth aged 16–18 years need to attend an educational program for at least 2 days a week. After that, youth must continue schooling until earning a diploma that provides a minimal entry level for the job market. The educational system consists of schools for different age groups, e.g., a differentiation is made in terms of elementary school and high school. Note that the Dutch educational system does not have a middle school or a junior high school. The majority of schools for different age groups are public schools and are financed by the government. All of them fall under the jurisdiction of the Dutch Inspection of Education.

Elementary school in the Netherlands consists of eight grades, with most children starting grade 3 at or around the age of 6. The Dutch school system acknowledges two types of schools: regular schools and schools for children with special needs. Children and youth can be enrolled in elementary or high schools for special education until they are 20 years old. Nationally, there are 4 clusters of schools for children who are in need of special education: schools providing education to children who are blind or have a visual impairment (cluster 1 schools; although the majority of children with visual impairments are enrolled in regular education programs); schools for children who are deaf or have an auditory impairment and children who have significant language impairments in their receptive or expressive (cluster 2 schools); schools for children who have a chronic physical illness, a physical handicap, or severe mental or cognitive disabilities (cluster 3 schools); and schools for children who have serious behavioral or psychiatric disorders (cluster 4 schools). Since the introduction of the law on Education that Fits (Passend Onderwijs, 2014), criteria for determining eligibility for special education or related services are set by the conglomerate of school boards within a region. Until this new law became effective, clusters 1, 2, and 3 all included a criterion directed by national law regarding a general intelligence score (e.g., 70 or lower for the cluster 3 and 70 or up for clusters 1 and 2). Other criteria for determining eligibility are related to testing receptive and expressive speech impairments, physical impairments, and psychopathology, depending on the cluster under study. Also, there are elementary schools enrolling only intellectually gifted children (i.e., so-called Leonardo schools or schools with Leonardo classes). To be enrolled in these Leonardo programs, children have to obtain a minimal general intelligence score of 130 points (i.e., 2 standard deviations above the mean), together with high problem-solving ability and creativity.

As shown here, in the context of selecting students for special programs, a general intelligence score (i.e., an estimate of g) has traditionally been used in the Netherlands, whereas less attention was paid to indices of intelligence (such as verbal intelligence) measured with a test. The COTAN states that because these overall intelligence scores are used for making important decisions (such as eligibility to special educational programs), the demands with respect to the psychometric properties of these test scores are more strict than those set for a test (or test scores) used for making less important decisions. Examples of tests used for making less important decisions are tests measuring a specific cognitive function, such as attention. Even with these higher standards, most intelligence tests used in the Dutch educational system are (psychometrically) appropriate for this predictive aim, when looking at the general intelligence scale. However, due to ceiling effects, certain tests are less suited for differentiating among scores at the top of the intelligence scores (≥120). Note, however, that, in practice, where intelligence criteria are being used, those responsible for the indication or admission often use the strict criterion of the
measured general intelligence score (or IQ score), ignoring the meaning of reliability intervals and issues of errors of measurement.

After elementary school is finished, children need to be enrolled in a high school program. As mentioned above, there are 4 clusters of special education programs, both at an elementary school level and at a high school level. Also, there are different streams of regular high school, with varying educational levels. The total duration varies as a function of the educational level, that is, a lower level of high school (VMBO, which combines vocational training with theory) takes 4 years, whereas the highest levels (HAVO and VWO, which prepare for the university for applied sciences or academic research universities respectively) consist of respectively 5-year or 6-year programs. From 2014–2015 on, all pupils enrolled in Grade 8 of a Dutch elementary school have to be assessed with an aptitude test, which is designed to help the teacher(s) to assess which level of high school would suit the individual pupil best. An intelligence test is not mandatory in this context, but intelligence tests are often administered to children enrolled in Grade 7 or Grade 8 to assist the teacher or a school team in predicting which level of high school would be most suitable for the individual child. However, professionals shall combine this information on intelligence with data obtained on other parameters, such as aptitude or achievement tests, to administer academic performances.

When the education staff or the parents are still in doubt at the end of Grade 8 of elementary school on what level of high school would suit the child best, they can choose to enroll the child in a high school orientation year (VMBO/HAVO or HAVO/VWO), once the child finishes elementary school. During this orientation year, children are taught (and tested on) materials at two adjacent levels of education. At the end of this year, the children need to be enrolled in one of these levels of high school. In some cases, intelligence tests are used in year 1 of high school to help predict which level of education would be most suitable. However, decisions are mostly made based on the academic performance.

Finally, after completing the levels VMBO or HAVO of high school, it is possible to be enrolled in the final years of the adjacent (next higher) level of high school (e.g., children who have passed the HAVO exams can be enrolled in Grade 5 of the VWO). In these cases, primarily the grade point averages obtained at the lower level of high school are used as an indication of whether an individual is capable of passing exams on a higher, adjacent level of high school.

The paragraphs above primarily discussed intelligence testing in the contexts of determining which educational program would suit best for the child or youngster. In more recent years, intelligence tests have increasingly been used for needs-based diagnostic purposes in the context of an evaluation of remediation, or to determine the child’s cognitive strengths and weaknesses to ascertain how to approach the child or what kind of remediation or intervention would suit the child best. Note, however, that intelligence test scores are only one aspect of the diagnostic process and should always be interpreted in the context of anamnesis, observations, and scores obtained on other tests, as well as within a transactional framework (Pameijer, 2009). Within a needs-based framework, intelligence assessment is only done inasmuch as it is relevant and needed for decision making on what is needed to meet the educational (as well as social emotional) needs of the particular child or youth, taking into account the context factors of family as well as the school, classroom, and teacher (Pameijer & Van Beukering, 2015).

Also, as the role of intelligence testing has traditionally been relatively large in the referral process for special education (until now schools for special elementary education have been using an intelligence criterion, as discussed above), teachers have learned to ask diagnosticians (psychologists and pedagogues) for intelligence testing for individual children. The move toward a more needs-based approach will hopefully lead to a more limited use of intelligence testing, limiting these assessments to situations in which the knowledge about the structure of intellectual capacities will contribute to intervention planning.

**Administrating intelligence tests in the Netherlands**

Since the mid 1950s, professionals with a university degree in psychology or pedagogy are responsible for administering intelligence tests and interpreting test results for children and youth (Amsing & De Beer, 2009). The degrees and professional credentials that are required for the administration and interpretation of these tests are defined in the manuals and professional guidelines set up by the Dutch Association of Psychologists (NIP; Sijtsma & Geertsema, 2010) and the guidelines set by the European Federation of Psychologists’ Associations (EFPA, 2012). The NIP guidelines state, for one, that a professional needs to have sufficient knowledge of tests, that is, the individual is able to select, administer, score, and interpret the tests. Professionals should always use instruments that have been approved by the COTAN (i.e., that the instrument has sufficient psychometric properties). If a test is not approved by the COTAN (or is lacking on one or more of the psychometric properties), then the professional has to justify why he or she has chosen to administer the test. In the professional code of ethics of the NIP, it is stated that psychologists should only accept and undertake
assessments autonomously, if they are sufficiently competent. However, the standards and ethical code are developed for psychologists and pedagogues with an academic degree in psychology or pedagogy who are members of the professional association. They are not formally included in governmental guidelines nor directed by law.

In contrast to the general description of credentials set by the NIP, the European Standards for test use set by the EFPA (2012) specify three levels of competence, when defining who is allowed to administer psychological tests and intelligence tests more specifically:

1. **EFPA Level 1**: those individuals who use specific tests in well-defined and constrained contexts, under the supervision of a more experienced test user, who is qualified at a higher level. The choice of tests and details of how they are used and applied are outside the person’s competence.
2. **EFPA Level 2**: those individuals that have a sufficient understanding of the technical psychometric qualities of the test for use, not for test construction, who can work independently as a test user in a specified and limited range of settings, and who has the necessary knowledge and skills to interpret a limited range of specific tests. Those individuals are not able to choose which test should be used nor are they able to interpret test scores beyond those based on the documentation provided for test users or provided in standard protocols. These are usually the junior psychologists, but also educators.
3. **EFPA level 3**: Those individuals who are specialized in testing and test use and who use tests as their core part of their practice. They are expected to have built a broad base of knowledge and skills in testing and test construction.

Note that both guidelines claim that all individuals have to act always in a professional and ethical manner, independent of their level of competence.

As the Netherlands since the 1980s no longer has a title protection for psychologists, nor specific certification requirements for professional practice, inappropriate administration and use of intelligence tests is a prominent risk. As the use of any test instrument—especially those measuring high stakes and especially in the case of youth—only has meaning and only can be interpreted in the transactional context, and with knowledge about cognitive development, there is a serious risk that use by insufficiently competent professionals, for example, teachers or psychological assistants, could lead to decisions that are detrimental to the child.

### Future directions for intelligence test development and use in the Netherlands

Traditional paper-and-pencil cognitive tests that are currently used provide only limited data on the processes underlying test performance and little understanding of the child’s potential to learn after obtaining instruction. They primarily quantify developed cognitive abilities and are “affected by many variables such as amount of education, test-taking skills, parental support and so on” (Grigorenko & Sternberg, 1998, p. 75). Alternatively, dynamic testing has been proposed as a promising method to quantify “the learning potential of the child during the acquisition of new cognitive operations” (Grigorenko & Sternberg, 1998, p. 75). In dynamic paradigms, children are continuously given feedback on their performance, while they are performing a sequence of tasks that become increasingly difficult or challenging for the child. The idea is that dynamic testing leads to testing the limits of an individual and “with graduated prompting and trial-by-trial-assessment this method could reveal the development of children’s strategy use while tested” (Resing, de Jong, Bosma, & Tunteler, 2009). Modern technology could provide an opportunity to construct computerized dynamic test protocols to measure learning potential and progress. Also, test items can be constructed using the availability of computers, Internet, and multimedia. In this context, guidelines on the do’s and don’ts as well as on how to assess the psychometric properties of computer-based testing (i.e., tests that are administered via the computer or on Internet) are still needed (Bartram & Hambleton, 2006). An example of an innovative approach to use computer-based testing in assessing children’s aptitude and cognitive abilities, and for training these functions by use of dynamic testing, is a Dutch computer program called Math Garden (Straatemeier, 2014), which aims to stimulate cognitive functions, learning, and academic skills in the context of dynamic testing. More than two thousand Dutch elementary schools participate in this project and, at the moment, more than 50 million tests have been administered in children, with more than 250.000 children having repeatedly used the computer program.

It is common practice to interpret (profile) scores on intelligence test batteries in terms of an individual’s strengths and weaknesses. Also, it is common for large amounts of subtest scatter, or differences between the highest and lowest scaled subtest scores, on an intelligence battery to be interpreted by professionals as an index for abnormality or cognitive impairment (Lezak, 2004), regardless of whether there is an abnormal level in the lowest scaled subtest scores (Binder & Binder, 2011). In 2013, Hurks et al. demonstrated, however, that subtest
scatter is highly common in a sample of healthy Dutch children aged 4.0–8 years, irrespective of the children’s ages. These results are in line with international studies reporting on test scatter in large samples of older children and healthy adults (Binder & Binder, 2011; Wisdom, Mignogna, & Collins, 2012). These results indicate that test scatter observed while administering intelligence scales is a characteristic of normal populations and therefore should not be automatically interpreted by a professional as an index for abnormality or cognitive impairment (Hurks et al., 2013).

Furthermore, it is highly relevant to study test development in the context of how fair intelligence tests are for minority groups in terms of the administration and interpretation of these measures (Kruse, 2015; Naglieri, 2015). Research has found that native Dutch children performed better on all subtests of an intelligence test compared to minority groups, even if similar ability levels are assumed (Wechsler, 2009). In 2012, Huijding, Hemker, and Van den Berg published guidelines for assessing fairness in testing in the Netherlands, as an addendum to the COTAN test reviewing system. Inasmuch as intelligence scores are being used as criteria for special educational support or admission to secondary education, extra care should be taken that the instruments used allow for equal opportunity for children of all ethnic groups and nationalities attending Dutch schools.

Finally, the question of who is actually qualified to perform intelligence assessments should be put on the agenda. Although the professional guidelines for psychologists and pedagogues are clear about competencies and qualifications, the fact that there is no national regulation leaves room for employers to hire less qualified (often cheaper) persons to perform assessments. Only then can the quality of the assessment and the relevance to needs-based education and intervention be guaranteed.

**About the authors**

Dr. Petra Hurks is an Associate Professor in the Faculty of Psychology and Neuroscience at Maastricht University, the Netherlands. She is the program director for the Bachelor program in Psychology at Maastricht University. Also, she currently serves as the Chair of the Dutch COTAN Evaluation system for test quality, which was founded by the Dutch Association of Psychologists. She is the co-author of the Dutch version of the Wechsler test for measuring intelligence in primary school and preschool children – third Edition (i.e., the WPPSI-III). Dr. Hurks’s major areas of scholarly interest concern: inter-individual differences in cognitive developmental trajectories and developing and evaluating tools to assess and treat (or stimulate) human cognitive abilities. She is on the editorial board of *Child Psychiatry & Human Development*.

She is program coordinator of the Master program in (Clinical) Child & Adolescent Psychology. She also is Chair of the Dutch Post-Master program for School Psychologists at RINO Amsterdam and the Radboud Center Social Sciences. Dr. Bakker is former President of the International School Psychology Association (ISPA) and currently member of the Board of Educational Affairs of the European Federation of Psychologists Associations (EFPA). Her areas of interest include ethics and professional development, children’s rights, evidence-based teaching and needs based assessment. Dr. Bakker is registered Child- & Adolescent Psychologist Specialist NIP, Health Psychologist and certified supervisor. She is also holder of the Europsy certificate in Education and Health.

**References**

Amsing, H. T. A., & De Beer, F. H. (2009). Selecting children with mental disabilities: A Dutch conflict over the demarcation of expertise in the 1950s. *Paedagogica Historica*, 45(1-2), 235–250.

Baldi, E. (2014). Geen struivogelpolitiek tegen de parasieten. Dirk Herderschée en het zwakzinnigenvraagstuk in het interbellum [Dirk Herderschée and the issue of mental retardation]. *Skript Historisch Tijdschrift*, 33(3), 145–158.

Bartram, D., & Hambleton, R. K. (2006). *Computer-based testing and the Internet. Issues and advances*. London, England: Wiley.

Binder, L. M., & Binder, A. L. (2011). Relative subtest scatter in the WAIS-IV standardization sample. *The Clinical Neuropsychologist*, 25(1), 62–71.

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic de niveau intellectual des anormaux. *L’Année Psychologique*, 11, 191–244.

Bos, A. & Magez, W. (2016). Het ene IQ is het andere niet: Van een psychometrische grootheid (IQ) naar een cognitief vaardighedsprefiel [One IQ is not the same as another IQ: From a psychometric unit (IQ) to a profile of cognitive skills]. *Tijdschrift voor Neuropsychologie*, 11(1), 12–15.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38(592), 10.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.

Cattell, R. B. (1971). *Intelligence: Its structure, growth and action*. Amsterdam, Netherlands: Elsevier.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the g-factor of intelligence. *Neuroimage*, 31, 1359–1365.

Depaepe, M. (1989). *Meten om Beter te Weten? Geschiedenis van de Experimenteel-Wetenschappelijke Richting in de Westerse Pedagogiek vanaf het Einde van de 19de Eeuw tot aan de Tweede Wereldoorlog* [Measuring to know more? The history of the experimental-scientific movement in Western pedagogy – from the end of the 19th century until World War II] (Unpublished doctoral dissertation). University of Leuven, Leuven, Belgium.

Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassing[ Test theory: Introduction to the psychological test theory and its applications]* (4th ed.). Houten, Netherlands: Bohn Stafleu van Loghum.
Education that fits. (2014). Retrieved from https://www.passenonderwijs.nl/overpassend onderwijs/in-het-kort/situatie-voor-passend-onderwijs/

European Federation of Psychologists' Associations Standards for Test Use. (2012). Performance requirements, context definitions and knowledge & skill specifications for the three EFPA levels of test user qualification. Brussels, Belgium: Author.

Egberink, I. J. L., Janssen, N. A. M., & Vermeulen, C. S. M. (2009 – 2014). COTAN Documentatie [COTAN Documentation]. (www.cotandocumentatie.nl). Amsterdam, Netherlands: Boom Test Uitgevers.

Evers, A. (2012). The internationalization of test reviewing: Trends, differences, and results. International Journal of Testing, 12(2), 136 – 156.

Fischer, G. H. (1974). Einführung in die Theorie Psychologischer Tests [Introduction to psychological test theory]. Bern: Huber.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). Essentials of cross-battery assessment (Vol. 84). Hoboken, NJ: John Wiley & Sons.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. Psychology Bulletin, 124(1), 75 – 111.

Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.) Handbook of developmental psychology. Englewood Cliffs, NJ: Prentice Hall.

Huijding, J., Hemker, B., & Van den Berg, R. (2012). Welke rol heeft de COTAN? Verantwoord en fair testgebruik [What is the role of the COTAN? Responsible and fair test use]. De Psycholoog, 4, 46 – 52.

Hurks, P. P. M., Hendriksen, J. G. M., Dek, J. E., & Kooij, A. P. (2013). Normal variability of children’s scaled scores on subtests of the Dutch Wechsler Preschool and Primary scale of Intelligence—Third Edition. The Clinical Neuropsychologist, 27(6), 988 – 1003.

Kaldenbach, Y. (2007). Hiërarchische analyse van de WISC-III subtests of the Dutch Wechsler Preschool and Primary scale of Intelligence—Third Edition. The Clinical Neuropsychologist, 27(6), 988 – 1003.

Kaldenbach, Y. (2006). De WISC-III anno 2006: Een voorstel van eenduidige en hiërarchische analyse, interpretatie en rapportage [The WISC in the year 2006: The proposal of an univocal and hierarchical analysis, interpretation and report]. Kind en Adolescent Praktijk, 5(3), 128 – 136.

Kaldenbach, Y. (2007). Hiërarchische analyse van de WISC-III nader toegelicht: Vragen en antwoorden [Hierarchical analyses of the WISC-III further explained: Questions and responses]. Kind en Adolescent Praktijk, 6(2), 72 – 80.

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman—Intelligentietest voor Adolescenten en Volwassenen (KAIT-NL). Handleiding. Nederlandse bewerking [Kaufman Intelligence test for adolescents and adults (KAIT-NL). Manual. Dutch adaptation]. Amsterdam: Pearson Information and Assessment BV.

Kaufman, J. C., Kaufman, S. B., & Plucker, J. A. (2013). Contemporary theories of intelligence. In D. Reisberg (Ed.), Oxford handbook of cognitive psychology. Oxford, England: Oxford University Press.

Knegtmans, P. J., & Kox, A. J. (2000). Tot nut en eer van de stad: Wetenschappelijk onderzoek aan de Universiteit van Amsterdam [Scientific research conducted at the University of Amsterdam]. Amsterdam, Netherlands: Amsterdam University Press.

Kruse, A. J. (in press). Cultural bias in testing: A review of literature and implications for music education. Applications of Research in Music Education.

Lezak, M. D. (2004). Neuropsychological assessment. New York, NY: Oxford University Press.

Luria, A. (1966). Higher cortical functions in man. New York, NY: Basic Books.

Matthews, G., Zeidner, M., & Roberts, R. D. (2004). Emotional intelligence: Science & myth. London, England: MIT Press.

Mulder, E., & Heyting, F. (1998). The Dutch curve: The introduction and reception of intelligence testing in the Netherlands, 1908 – 1940. Journal of the History of the Behavioral Sciences, 34(4), 349 – 366.

Naglieri, J. A. (2015). Hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In Handbook of intelligence (pp. 295–316). New York, NY: Springer.

Naglieri, J. A., & Das, J. P. (1997). Das-Naglieri cognitive assessment system. Itasca, IL: Riverside.

Pameijer, N. (2009). Handelingsgerichte diagnostiek. In De Wet op de jeugdzorg in de dagelijkse praktijk (pp. 159 – 192). Houten, Belgium: Bohn Staflue van Loghum.

Pameijer, N., & Van Beukering, T. (2015). Handelingsgerichte diagnostiek in het onderwijs: Een praktijkmodel voor diagnostiek en advisering [Needs-based diagnostics in education: A model for diagnostics and advice]. Leuven, Belgium: Uitgeverij Acco.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Nielsen & Lydiche.

Resing, W. C. M. (2015). Handboek intelligentietheorie en testgebruik [Handbook Intelligence: Theory and Test use]. Amsterdam, Netherlands: Pearson Assessment and Information B.V.

Resing, W. C. M., Bleichrodt, N., Drenth, P. J. D., & Zaal, J. N. (2012). RAKIT-2 Revisie Amsterdamse Kinder Intelligentie Test. [Revision Amsterdam intelligence test for children]. Amsterdam, the Netherlands: Pearson Assessment and Information B.V.

Resing, W., de Jong, F. M., Bosma, T., & Tunteler, E. (2009). Learning during dynamic testing: Variability in strategy use by indigenous and ethnic minority children. Journal of Cognitive Education and Psychology, 8(1), 22 – 37.

Sijtsma, K., & Geertsena, H. (2010). Algemene standaard testgebruik [General Standard Test use]. Utrecht, Netherlands: Nederlands Instituut voor Psychologen (NIP).

Straatemeier, M. (2014). Math garden: A new educational and scientific instrument. Amsterdam, Netherlands: University of Amsterdam.

Tellegen, P. J., & Laros, J. (2011). SON-R 6-40 Snijders-Oomen Niet-verbale intelligentietest. Verantwoording. [Snijders-Oomen non-verbal intelligence test]. Amsterdam, Netherlands: Hogrefe.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). Snijders-Oomen Niet-verbale Intelligentietest SON-R 2.5-7. Handleiding en verantwoording [Snijders-Oomen non-verbal intelligence test SON-R 2.5-7. Administration and technical manual]. Lisse, Netherlands: Swets & Zeitlinger.

Thurstone, L. L. (1938). Primary mental abilities. Chicago, IL: University of Chicago Press.

Van Batenburg, Th. A., & Van der Werf, M. P. C. (2004). NSCT; Niet Schoolse Cognitieve Capaciteiten Test voor groep 4, 6 en 8 in het basisonderwijs. Verantwoording, normering en handleiding. Groningen, Netherlands: GION.
Vegt, E. R., & Metselaar, E. E. (2011). Handleiding ADIT Adaptieve Digitale Intelligentie Test [Manual ADIT: Adaptive Digital Intelligence Test]. Apeldoorn, Netherlands: A-Vision Holding.

Verweij, A. C. (2002). Instaptoets Voorgezet Onderwijs (IVO) [IVO Test for high school students. Manual]. Handleiding. Apeldoorn, Netherlands: Stichting Verweij.

Verweij, A. C. (2008). Testserie voor Plaatsing in het Voortgezet Onderwijs, Instaptoets Voortgezet Onderwijs (IVO). Internetversie. Handleiding [IVO Test for high school students. An Internet version]. Ugcelen, Netherlands: A-VISION.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1944). The measurement of adult intelligence (3rd ed.). Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1946). Wechsler-Bellevue Intelligence Scale, Form II. Manual for administering and scoring this test. New York, NY: Psychological Corporation.

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. American Psychologist, 30, 135–139.

Wechsler, D. (2009). WPPSI-III_NL. Nederlandse bewerking, Technische handleiding. Amsterdam, Netherlands: Pearson Assessment and Information BV.

Wisdom, N. M., Mignogna, J., & Collins, R. L. (2012). Variability in Wechsler Adult Intelligence Scale-IV subtest performance across age. Archives of Clinical Neuropsychology, 27(4), 389–397.