INTRODUCTION

Traditional tests for general intelligence like the Stanford-Binet and the Wechsler intelligence tests have been criticized on the point that these tests measure the end result of prior learning rather than learning potential. By merely reflecting the end result of prior learning general intelligence tests would underestimate the learning ability of persons who have had fewer opportunities to acquire the knowledge and skills to perform well in a test situation. In particular, members of ethnic minorities, persons from lower socio-economic background and persons with learning problems would be at a disadvantage when tested with a traditional general intelligence test.

Traditional intelligence tests have also been criticized on the basis of their contents by advocates of culture fair intelligence tests. Because the traditional tests often make an appeal to specific language skills, both in test contents and instructions, these tests would place members of cultural minority groups at a disadvantage. This argument also applies to persons with hearing-, speech- and language problems. For all these groups, low performance on the test might primarily reflect poor verbal knowledge instead of poor reasoning or learning ability. This criticism has led to the development of nonverbal intelligence tests which aim at minimizing the reliance on acquired knowledge and verbal ability, such as Raven's Progressive Matrices (Raven, 1938), the SON-tests (Snijders-Oomen, 1943) and Cattell's Culture Fair Intelligence Test (Cattell, 1950).

In the early forties, Snijders-Oomen (1943) constructed a nonverbal intelligence scale (SON) intended for the assessment of deaf children. Intelligence was defined by her in terms of learning ability; the extent to which children could profit from instruction at school. The SON-test developed by Snijders-Oomen was the first test that covered a wide area of intelligence without being dependent on the use of language. The scale has been revised several times and is especially suited for the intelligence assessment of immigrant children, children from cultural minorities and children with hearing-, speech- and language problems.

In this paper the latest revisions of the SON-test, the SON-R 2,5-7 and the SON-R 5,5-17, will be described. After a short summary of the history of the SON-tests and the description of the tests, some research results with immigrant subjects will be reviewed. In the discussion special attention will be given to the similarities and differences of the SON-R compared to general intelligence tests and learning potential tests.
HISTORY OF THE SON-TESTS

In her work with children at an institute for the deaf Snijders-Oomen was confronted with problems of assessing the learning ability of children who were severely handicapped in their language development. General intelligence tests were not suited for this purpose due to reliance on verbal skills, while nonverbal tests at that time consisted mainly of performance tests related to spatial abilities (like mazes, form boards, mosaics). After extensive experimentation with existing and newly developed tasks she constructed a test series which also included nonverbal subtests related to abstract and concrete reasoning. Capacities for abstraction and combination were considered especially important for the ability to participate in the educational system (Snijders-Oomen, 1943, pp. 25-28). Mental age norms were constructed for deaf children from 4 to 14 years of age.

In the subsequent revision the test series was expanded and standardized for deaf and hearing children from 3 to 17 years (Snijders & Snijders-Oomen, 1970). The latest revision for the older children, published in 1988, is the SON-R 5,5-17 (Snijders, Tellegen & Laros, 1989; Laros & Tellegen, 1991). In 1998 the SON-R 2,5-7, a new revision of the test for the pre-school age group was be published (Tellegen, Winkel, Wijnberg-Williams & Laros, 1998). Common to the revisions of the SON-tests is the primary goal to examine a broad spectrum of intelligence without being dependent on language. Due to the nonverbal character of the SON-tests, the test materials can be used internationally without modifications; the manuals, scoring forms and computer program of the SON-R tests have been published in the English, German and Dutch languages. In 2005 a standardisation study for the SON-R 2,5-7 has been completed in Germany while standardisations will start this year in Great Britain, the Czech Republic and in Brazil. Research for a new revision of the SON-R 5,5-17 has started in 2004. The new test, with working title SON-E 6-60, is expected to be published in 2008. Several studies are performed in other countries, for instance China, Thailand, Indonesia, Kenya, Morocco, Brazil and Peru to increase the general application of the pictures that are used in the test (Tellegen & Laros, 2004). Below, some pictures are presented from the research in Nakuru, Kenya.
SUMMARY OF CHARACTERISTICS

In table 1 an overview is presented of some important characteristics of the two SON-R tests: the SON-R 2,5-7 for young children and the SON-R 5,5-17 for the older children.

Table 1
Some characteristics of the SON-R tests

<table>
<thead>
<tr>
<th></th>
<th>SON-R 2,5-7</th>
<th>SON-R 5,5-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range</td>
<td>2;6 – 6;11 years</td>
<td>5;6 – 6;11 years</td>
</tr>
<tr>
<td>Number of subtests</td>
<td>6 subtests</td>
<td>7 subtests</td>
</tr>
<tr>
<td>Administration</td>
<td>individually</td>
<td>individually</td>
</tr>
<tr>
<td>Duration</td>
<td>50 minutes</td>
<td>90 minutes</td>
</tr>
<tr>
<td>Sample size standardisation</td>
<td>N=1.124</td>
<td>N=1.350</td>
</tr>
<tr>
<td>Reliability</td>
<td>.90</td>
<td>.93</td>
</tr>
<tr>
<td>Generalisability</td>
<td>.78</td>
<td>.85</td>
</tr>
</tbody>
</table>

On the website [www.testresearch.nl](http://www.testresearch.nl) a lot of information on the SON-tests can be found. On the website are different sections in the English, German and Dutch language. The website can also be used to update the computer program that goes with the test.

Also, on the website, links are provided with the publisher of the test, Hogrefe Verlag in Germany, and with different distributors, like Testzentrale (Germany), The Test Agency (Great Britain), Hans Huber Verlag (Switzerland), Testcentrum Praha (Czech Republic) and Boom test uitgevers (The Netherlands and Belgium).
Evaluation

In The Netherlands, all psychological tests are evaluated by the test commission (COTAN) of the Dutch Institute of Psychologists (NIP). On seven aspects tests are evaluated as 'insufficient', 'sufficient' or 'good' (see table 2). On all aspects both the SON-tests were rated as good which is rather exceptional. The SON-tests also belong to the few tests that have been approved by the Dutch government to be used in important decisionmaking situations.

Table 2
Evaluation by the Dutch test Commission (COTAN)

<table>
<thead>
<tr>
<th></th>
<th>SON-R 2,5-7</th>
<th>SON-R 5,5-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Materials</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Manual</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Norms</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Reliability</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Construct validity</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>good</td>
<td>good</td>
</tr>
</tbody>
</table>

DESCRIPTION OF THE SON-R 2,5-7

The SON-R 2,5-7 is a general intelligence test for young children. The test assesses a broad spectrum of cognitive abilities without involving the use of language. This makes it especially suitable for children who have problems or handicaps in language, speech or communication, for instance, children with a language, speech or hearing disorder, deaf children, autistic children, children with problems in social development, and immigrant children with a different native language.

A number of features make the test particularly suitable for less gifted children and children who are difficult to test. The materials are attractive, the tasks diverse. The child is given the chance to be active. Extensive examples are provided. Help is available on incorrect responses, and the discontinuation rules restrict the administration of items that are too difficult for the child.

The SON-R 2,5-7 differs in various aspects from the more traditional intelligence tests, in content as well as in manner of administration. Therefore, this test can well be administered as a second test in cases where important decisions have to be taken, on the basis of the outcome of a test, or if the validity of the first test is in doubt.

Although the reasoning tests in the SON-R 2,5-7 are an important addition to the typical performance tests, the nonverbal character of the SON tests limits the range of cognitive abilities that can be tested. Other tests will be required to gain an insight into verbal development and abilities. However, for those groups of children for whom the SON-R 2,5-7 has been specifically designed, a clear distinction must be made between intelligence and verbal development.
The subtests of the SON-R 2,5-7

The SON-R 2,5-7 is composed of six subtests:
1. Mosaics,
2. Categories,
3. Puzzles,
4. Analogies,
5. Situations and
6. Patterns.

The subtests are administered in this sequence. The tests can be grouped into two types: reasoning tests (Categories, Analogies and Situations) and more spatial, performance tests (Mosaics, Puzzles and Patterns). The six subtests consist, on average, of 15 items of increasing difficulty. Each subtest consists of two parts that differ in materials and/or directions. In the first part the examples are included in the items. The second part of each subtest, except in the case of the Patterns subtest, is preceded by an example, and the subsequent items are completed independently.

**Mosaics**
The subtest Mosaics consists of 15 items. In Mosaics, part I, the child is required to copy several simple mosaic patterns in a frame using three to five red squares. The level of difficulty is determined by the number of squares to be used and whether or not the examiner first demonstrates the item.

In Mosaics II, diverse mosaic patterns have to be copied in a frame using red, yellow and red/yellow squares. In the easiest items of part II, only red and yellow squares are used, and the pattern is printed in the actual size. In the most difficult items, all of the squares are used and the pattern is scaled down.

**Categories**
Categories consists of 15 items. In Categories I, four or six cards have to be sorted into two groups according to the category to which they belong. In the first few items, the drawings on the cards belonging to the same category strongly resemble each other. For example, a shoe or a flower is shown in different positions. In the last items of part I, the child must him or herself identify the concept underlying the category: for example, vehicles with or without an engine.
Categories II is a multiple choice test. In this part, the child is shown three pictures of objects that have something in common. Two more pictures that have the same thing in common have then to be chosen from another column of five pictures. The level of difficulty is determined by the level of abstraction of the shared characteristic.

**Puzzles**
The subtest Puzzles consists of 14 items. In part I, puzzle pieces must be laid in a frame to resemble the given example. Each puzzle has three pieces. The first few puzzles are first demonstrated by the examiner. The most difficult puzzles in part I have to be solved independently.

In Puzzles II, a whole must be formed from three to six separate puzzle pieces. No directions are given as to what the puzzles should represent; no example or frame is used. The number of puzzle pieces partially determines the level of difficulty.
Analogies
The subtest Analogies consists of 17 items. In Analogies I, the child is required to sort three, four or five blocks into two compartments on the basis of either form, color or size. The child must discover the sorting principle him or herself on the basis of an example. In the first few items, the blocks to be sorted are the same as those pictured in the test booklet. In the last items of part I, the child must discover the underlying principle independently: for example, large versus small blocks.

Analogies II is a multiple choice test. Each item consists of an example-analogy in which a geometric figure changes in one or more aspect(s) to form another geometric figure. The examiner demonstrates a similar analogy, using the same principle of change. Together with the child, the examiner chooses the correct alternative from several possibilities. Then, the child has to apply the same principle of change to solve another analogy independently. The level of difficulty of the items is related to the number and complexity of the transformations.

Situations
The subtest Situations consists of 14 items. Situations I consists of items in which one half of each of four pictures is shown in the test booklet. The child has to place the missing halves beside the correct pictures. The first item is printed in color in order to make the principle clear. The level of difficulty is determined by the degree of similarity between the different halves belonging to an item.
Situations II is a multiple choice test. Each item consists of a drawing of a situation with one or two pieces missing. The correct piece (or pieces) must be chosen from a number of alternatives to make the situation logically consistent. The number of missing pieces determines the level of difficulty.

Patterns
The subtest Patterns consists of 16 items. In this subtest the child is required to copy an example. The first items are drawn freely, then pre-printed dots have to be connected to make the pattern resemble the example. The items of Patterns I are first demonstrated by the examiner and consist of no more than five dots.

The items in Patterns II consist of five, nine or sixteen dots and have to be copied by the child without help. The level of difficulty is determined by the number of dots and whether or not the dots are pictured in the example pattern.
Reasoning tests, spatial tests and performance tests

Reasoning tests
Reasoning abilities have traditionally been seen as the basis for intelligent functioning (Carroll, 1993). Reasoning tests form the core of most intelligence tests. They can be divided into abstract and concrete reasoning tests. Abstract reasoning tests, such as Analogies and Categories, are based on relationships between concepts that are abstract, i.e., not bound by time or place. In abstract reasoning tests, a principle of order must be derived from the test materials presented, and applied to new materials. In concrete reasoning tests, like Situations, the object is to bring about a realistic time-space connection between persons or objects (see Snijders, Tellegen & Laros, 1989).

Spatial tests
Spatial tests correspond to concrete reasoning tests in that, in both cases, a relationship within a spatial whole must be constructed. The difference lies in the fact that concrete reasoning tests concern a meaningful relationship between parts of a picture, and spatial tests concern a ‘form’ relationship between pieces or parts of a figure (see Snijders, Tellegen & Laros, 1989; Carroll, 1993). Spatial tests have long been integral components of intelligence tests. The spatial subtests included in the SON-R 2,5-7 are Mosaics and Patterns. The subtest Puzzles is more difficult to classify, as the relationship between the parts concerns form as well as meaning. We expected the performance on Puzzles and Situations to relate to concrete reasoning ability. However, the correlations and factor analysis show that Puzzles is more closely associated with Mosaics and Patterns.

Performance tests
An important characteristic that Puzzles, Mosaics and Patterns have in common is that the item is solved while manipulating the test stimuli. That is why these three subtests are called performance tests. In the three reasoning tests (Situations, Categories and Analogies), in contrast, the correct solution has to be chosen from a number of alternatives. For the rest, the six subtests are very similar in that perceptual and spatial aspects as well as reasoning ability play a role in all of them.

The performance subtests of the SON-R 2,5-7 can be found in a similar form in other intelligence tests. However, only verbal directions are given in these tests. Reasoning tests can also regularly be found in other intelligence tests, but then they often have a verbal form (such as verbal analogies).

Characteristics of administration of the SON-R 2,5-7

Individual intelligence test
Most intelligence tests for children are administered individually. The SON-R 2,5-7 follows this tradition for the following reasons:

• the directions can be given nonverbally,
• feedback can be given in the correct manner,
• testing can be tailored to the level of each individual child,
• the examiner can encourage children who are not very motivated or cannot concentrate; personal contact between the child and the examiner is essential for effective testing, certainly for children up to the age of four to five years.
**Nonverbal intelligence test**

The SON-R 2,5-7 is nonverbal. This means that the test can be administered without the use of spoken or written language. The examiner and the child are not required to speak or write and the testing materials have no language component. One is, however, allowed to speak during the test administration, otherwise an unnatural situation would arise. The manner of administration of the test depends on the communication abilities of the child. The directions can be given verbally, nonverbally with gestures or using a combination of both. Care must be taken when giving verbal directions that no extra information is given.

No knowledge of a specific language is required to solve the items being presented. However, level of language development, for example, being able to name objects, characteristics and concepts, can influence the ability to solve the problems correctly. Therefore the SON-R 2,5-7 should be considered a nonverbal test for intelligence rather than a test for nonverbal intelligence.

**Directions**

An important part of the directions to the child is the demonstration of (part of) the solution to a problem. An example item is included in the administration of the first item on each subtest, and detailed directions are given for all first items. Once the child understands the nature of the task, the examiner can shorten the directions for the following items. If the child does not understand the directions, they can be repeated. In the second part of each subtest an example is given in advance. Once the child understands this example, he or she can do the following items independently.

**Feedback**

The examiner gives feedback after each item. In the SON-R 5,5-17, feedback is limited to telling the child whether his or her answer is correct or incorrect. In the SON-R 2,5-7 the examiner indicates whether the solution is correct or incorrect, and, if the answer is incorrect, he/she also demonstrates the correct solution for the child. The examiner tries to involve the child when correcting the answer, for instance, by letting him or her perform the last action. However, the examiner does not explain why the answer was incorrect. By giving feedback, a more normal interaction between the examiner and the child occurs, and the child gains a clearer understanding of the task. The child is given the opportunity to learn and to correct him or herself. In this respect a similarity exists between the SON-tests and tests for learning potential (Tellegen & Laros, 1993a).

**Entry procedure and discontinuation rule**

Each subtest begins with an entry procedure. Based on age and, when possible, the estimated cognitive level of the child, a start is made with the first, third or fifth item. This procedure was chosen to prevent children from becoming de-motivated by being required to solve too many items that are below their level. The design of the entry procedure ensures that the first items the child skips would have been solved correctly. Should the level chosen later appear to be too difficult, the examiner can return to a lower level. However, because of the manner in which the test has been constructed, this should occur infrequently.

Each subtest has rules for discontinuation. A subtest is discontinued when a total of three items has been incorrectly solved. The mistakes do not have to be consecutive. The three performance subtests are also discontinued when two consecutive mistakes are made in the second part. Frequent failure often has a drastically de-motivating effect on children and can result in refusal to go on.
**Time factor**
The speed with which the problems are solved plays a very subordinate role in the SON-R 2,5-7. A time limit for completing the items is used only in the second part of the performance tests. The time limit is generous. Its goal is to allow the examiner to end the item. The construction research showed that children who go beyond the time limit are seldom able to find a correct solution when given more time.

**Duration of test administration**
The administration of the SON-R 2,5-7 takes about 50 minutes (excluding any short breaks during administration). During the standardization research the administration took between forty and sixty minutes in 60% of the cases. For children with a specific handicap, the administration takes about five minutes longer. For children two years of age, administration time is shorter; nearly 50% of the two-year-olds complete the test in less than forty minutes.

**Standardization**
The SON-R 2,5-7 is meant primarily for children in the age range from 2;6 to 7;0 years. The norms were constructed using a mathematical model in which performance is described as a continuous function of age. An estimate is made of the development of performance in the population, on the basis of the results of the norm groups. These norms run from 2;0 to 8;0 years. In the age group from 2;0 to 2;6 years, the test should only be used for experimental purposes. In many cases the test is too difficult for children younger than 2;6 years. Often, they are not motivated or concentrated enough to do the test. However, in the age group from 7;0 to 8;0 years, the test is eminently suitable for children with a cognitive delay or who are difficult to test. The easy starting level and the help and feedback given can benefit these children. For children of seven years old who are developing normally, the SON-R 5,5-17 is generally more appropriate.

The scaled subtest scores are presented as standard scores with a mean of 10 and a standard deviation of 3. The scores range from 1 to 19. The SON-IQ, based on the sum of the scaled subtest scores, has a mean of 100 and a standard deviation of 15. The SON-IQ ranges from 50 to 150. Separate total scores can be calculated for the three performance tests (SON-PS) and the three reasoning tests (SON-RS). These have the same distribution characteristics as the IQ score. When using the computer program, the scaled scores are based on the exact age; in the norm tables age groups of one month are presented. With the computer program, a scaled total score can be calculated for any combination of subtests.

In addition to the scaled scores, based on a comparison with the population of children of the same age, a reference age can be determined for the subtest scores and the total scores. This shows the age at which 50% of the children in the norm population perform better, and 50% perform worse. The reference age ranges from 2;0 to 8;0 years. It provides a different framework for the interpretation of the test results, and can be useful when reporting to persons who are not familiar with the characteristics of deviation scores. The reference age also makes it possible to interpret the performance of older children or adults with a cognitive delay, for whom administration of a test, standardized for their age, is practically impossible and not meaningful.

As with the SON-R 5,5-17, no separate norms for deaf children were developed for the SON-R 2,5-7. Our basic assumption is that separate norms for specific groups are only required when a test discriminates against a special group of children because of its contents or the manner in which it is administered. Research using the SON-R 2,5-7 and the SON-R 5,5-17 with deaf children shows that this is absolutely not the case for deaf children with the SON tests.
Performance of immigrant children on the SON-R 2,5-7

In this section results are presented of the test performances of children one or both of whose parents were born outside the Netherlands. These children were tested in the standardization research (N=147), or attended a preschool playgroup (N=8), or a primary school where complementary research projects were carried out (N=54). Of these 209 children, 118 were immigrant children (both parents were born outside the Netherlands) and the remaining 91 children belonged to the mixed group (one parent born outside the Netherlands). Later, the results of the immigrant children will be compared with the results of 90 children participating in OPSTAP(JE), a program to stimulate the development of immigrant children.

The test scores of the mixed and immigrant groups were compared with the scores of the native Dutch children from the standardization research. The mean ages were 4;9 years in the native Dutch group, 5;3 years in the mixed group and 5;6 years in the immigrant group.

The mean scores are presented in the following table. The performances of the children with a mixed background differed only slightly from the performances of native Dutch children. The differences for the total scores were negligible, and none of the differences in the subtest scores were significant at the 5% level. However, the mean scores of the immigrant children were clearly lower than those of the native Dutch children. The mean IQ score of the immigrant children was nearly 8 points lower than that of the native Dutch children. With the exception of Analogies, all differences were significant at the 1% level.

In the group of immigrant children the differences between the mean subtest scores were slight. The results show that the lower performances of the immigrant children were not caused or worsened specifically by the subtests Categories, Situations and Puzzles. These subtests use meaningful picture materials and might therefore have a culture specific meaning. The mean score on these three subtests was equal to the mean score on Patterns, Mosaics and Analogies. These last three subtests use nonmeaningful picture materials such as geometrical forms. No differences were found between the mean scores on the Performance Scale and the Reasoning Scale in the immigrant or in the mixed group.

Table 3
Test Scores of Native Dutch Children, Immigrant Children and Children of Mixed Parentage

<table>
<thead>
<tr>
<th>Score</th>
<th>Native Dutch (N=969)</th>
<th>Mixed (N=91)</th>
<th>Immigrant (N=118)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SON-PS</td>
<td>100.7 (15.2)</td>
<td>100.2 (14.3)</td>
<td>93.4 (13.6)</td>
</tr>
<tr>
<td>SON-RS</td>
<td>100.4 (14.8)</td>
<td>100.5 (15.2)</td>
<td>93.8 (15.6)</td>
</tr>
<tr>
<td>SON-IQ</td>
<td>100.7 (14.9)</td>
<td>100.6 (15.2)</td>
<td>92.8 (14.4)</td>
</tr>
</tbody>
</table>

Relationship with SES level

Information about the level of education and occupation of the parents was available for most children. The SES index, calculated on the basis of these data, had a mean of 5.1 (sd=2.8) in the mixed group, a mean of 2.5 (sd=2.7) in the immigrant group, and a mean of 4.9 in the native Dutch group (sd=2.5). The SES index of the immigrant children was significantly (p<.01) lower than the SES index of the native Dutch children.
In the table below the percentage of children at each SES level is presented for each group. The distribution curve of the mixed group was slightly flatter than that of the native Dutch group; the distribution of the immigrant group was very skewed. In comparison to the native Dutch group, more than three times as many children of the immigrant group had a low SES level, whereas the number of children with a high SES level in the native Dutch group was more than three times as high as that in the immigrant group. The mean IQ scores for each SES level are also presented in the table. Within each group a clear and comparable relationship existed between SES level and IQ, and no significant interaction effect was found. When the SES level was controlled, the differences among the three groups almost disappeared and were no longer significant. The difference of nearly eight IQ points between the immigrant and the native Dutch group decreased, after controlling for the SES level, to three points.

Table 4
Relationship Between Group, SES Level and IQ

<table>
<thead>
<tr>
<th>SES Level</th>
<th>Native Dutch (N=963)</th>
<th>Mixed (N=90)</th>
<th>Immigrant (N=117)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>18% 92.9 (14.9)</td>
<td>21% 94.3 (11.9)</td>
<td>61% 90.3 (13.5)</td>
</tr>
<tr>
<td>Below average</td>
<td>33% 98.8 (13.9)</td>
<td>27% 98.3 (16.0)</td>
<td>22% 94.6 (12.3)</td>
</tr>
<tr>
<td>Above average</td>
<td>31% 102.9 (13.8)</td>
<td>28% 102.7 (15.9)</td>
<td>11% 99.1 (17.8)</td>
</tr>
<tr>
<td>High</td>
<td>19% 108.1 (14.9)</td>
<td>24% 106.5 (14.2)</td>
<td>6% 102.9 (16.6)</td>
</tr>
</tbody>
</table>

Differentiation according to country of birth
The largest immigrant groups in the Netherlands come from Surinam, the Antilles, Morocco and Turkey. These groups are most strongly represented in this research group. Children of parents born in Surinam, Morocco and Turkey had mean scores close to 90. The small group of children from other African and South American countries had the same mean score. The Antillean and Asian children had scores close to 100 and the small group of children from other Western countries performed above average. For children with one parent born outside the Netherlands, the differences in mean IQ scores were slight.

Comparison with other tests
The mean IQ score of the 118 immigrant children in this research project with the SON-R 2,5-7 was 92.8, nearly 2 points higher than the mean IQ score of 91.0 of the immigrant children participating in the standardization research of the SON-R 5,5-17 (N=61). In comparison to the SON-R 5,5-17, the scores of the Turkish children were higher, while the scores of the Surinam/Antillean children were lower.

Research was done with the RAKIT in different immigrant groups by Resing, Bleichrodt and Drenth (1986). The RAKIT is an intelligence test with verbal and performance tasks for children 4 years and older. In the age group of 5;8 years the mean RAKIT IQ in the Surinam/Antillean group was 89.6; in the Turkish group this was 80.0 and in the Moroccan group 80.5. Each group consisted of approximately 60 children. The mean IQ scores on the SON-R 2,5-7 in the three ethnic groups were 3, 9 and 11 points higher respectively. Using the LEM (Learning test for Ethnic Minorities; Hamers, Hessels & van Luit, 1991), research was done with Turkish and Moroccan children five and
six years of age. The LEM was specially designed to measure learning potential and to depend as little as possible on culture specific knowledge and skills. The Turkish and Moroccan groups consisted of 120 children each. The mean standardized total scores of the Turkish and the Moroccan children were 83.5 and 84.4 respectively. This means that their mean score on the LEM was approximately 6 points lower than the mean IQ score of Turkish and Moroccan children on the SON-R 2,5-7.

The conclusion on the basis of these comparisons is that immigrant children get better results on the SON-R 2,5-7 than on the RAKIT and the LEM. Comparisons of the SON-R 2,5-7 and other tests, administered to the same children, indicate also that the SON-R 2,5-7 is much less dependent on culture specific knowledge and skills (see Tellegen et al., 1998, section 9.11).

**The test performances of children participating in OPSTAP(JE)**

OPSTAP is a family intervention program for immigrant families and has been used in the Netherlands since 1987. It is the Dutch version of the program HIPPY (Home Intervention Programme for Preschool Youngsters; Lombard, 1981) that was developed in Israel. OPSTAP is aimed at helping mothers of immigrant children in the kindergarten age range. OPSTAPJE has been operating for a few years now and is aimed at helping mothers with children in the preschool age range. The goal of the programs is to enhance the mother’s ability to stimulate the child in his or her development. This is achieved by (group)discussions, by providing materials and by supplying exercises for the child.

In 1994, research using the SON-R 2,5-7 was carried out with a number of children who were participating in OPSTAP or OPSTAPJE. In general, the test was administered at the end of the two-year intervention period. Three of the four examiners (all of them female) had participated in OPSTAP(JE) as coordinator or trainee. One examiner was of Moroccan descent and one of Surinam descent. A total of 105 OPSTAP(JE) children were tested. We have limited the presentation of the results to those Surinam, Moroccan and Turkish children, whose parents were both born outside the Netherlands (N=90). A good comparison can be made between these groups and similar immigrant groups discussed previously that have not, as far as we know, participated in an intervention program.

The percentage of boys in both the OPSTAP(JE) group and the immigrant comparison group was 53%. The age varied from two to seven years and had a mean of 5;0 years (sd=1;4 years). The number of Surinam children was 33; the number of Moroccan children was 22 and the number of Turkish children was 35.

In the next table the mean scores are presented of the OPSTAP(JE) children, of the immigrant children from the comparison group and of the native Dutch children from the standardization research. The mean score of the 90 OPSTAP(JE) children was 102.8. This was two points higher than the mean score of the native Dutch children. However, the difference was not significant. The mean score of the OPSTAP(JE) children was 12.5 points higher than that of the immigrant children from the comparison group. The difference according to country of birth was largest for Moroccan children and least for Surinam children. A variance analysis carried out with country of birth and participation in OPSTAP(JE) as factors, showed that neither the interaction effect nor the main effect for country of birth was significant. However, the main effect for participation in OPSTAP(JE) was highly significant (F[1,167]=33.77, p<.01).
Table 5
Mean IQ scores of Surinam, Turkish and Moroccan children who had participated in the OPSTAP(JE) project

<table>
<thead>
<tr>
<th>Country of birth of parents</th>
<th>OPSTAP (JE) Immigrant</th>
<th>Comparison Group</th>
<th>Native Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>N</td>
</tr>
<tr>
<td>Surinam</td>
<td>33</td>
<td>98.4 (17.6)</td>
<td>36</td>
</tr>
<tr>
<td>Morocco</td>
<td>22</td>
<td>106.5 (11.3)</td>
<td>25</td>
</tr>
<tr>
<td>Turkey</td>
<td>35</td>
<td>104.7 (11.1)</td>
<td>22</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>102.8 (14.2)</td>
<td>83</td>
</tr>
</tbody>
</table>

The possibility exists that factors other than participation in OPSTAP(JE) contributed to these differences, such as, for instance, the SES level of the parents. A selection effect may have occurred in the decision for parents to participate in OPSTAP(JE), or when parents agreed to participate in this research. Another difference is that the test was administered at home in the OPSTAP(JE) research, and at school in most of the other research projects. The ethnic background of the examiners appeared to have had no influence. The scores of the children who were tested by the two immigrant examiners were on average two points lower than the scores of the children who were tested by the two Dutch examiners. Furthermore, the scores of the children who were tested by an examiner from their own ethnic group were no higher than those of the other children. What could, of course, have played a role is, that all four examiners had a great deal of experience with immigrant children and were therefore well able to motivate and stimulate the children. In order to be able to give an unambiguous evaluation of the effect of OPSTAP(JE), research needs to be done with a pretest, post-test, control group design, with the examiner as variable to be controlled for.
THE SON-R 5,5-17

Composition of the SON-R 5,5-17

In sequence of administration the test series consists of the following 7 subtests:
1. Categories,
2. Mosaics,
3. Hidden Pictures,
4. Patterns,
5. Situations,
6. Analogies, and
7. Stories.

Categories
The subject is shown three drawings of objects or situations that have something in common. The subject has to discover the concept underlying the three pictures and is required to choose, from five alternatives, those two drawings which depict the same concept. The difficulty of the items is related to the degree of abstraction of the underlying concept. For example, in an easy item the concept is 'fruit' and in one of the most difficult items the concept is 'art'.
**Mosaics**
Various mosaic patterns, presented in a booklet, have to be copied by the subject using nine red/white squares. There are six different sorts of squares. With the easy items, only two sorts are used while all six sorts are used with the difficult items.

**Hidden Pictures**
A certain search object (for instance a kite) is hidden fifteen times in a drawing. The size and the position of the hidden object varies. After focusing on the search object, the subject has to indicate the places where it is hidden.
**Patterns**
In the middle of a repeating pattern of one or two lines a part is left out. The subject has to draw the missing part of the lines in such a way that the pattern is repeated in a consistent way. The difficulty of the items is related to the number of lines, the complexity of the line pattern and the size of the missing part.

**Situations**
The subject is shown a picture of a concrete situation in which one or more parts are missing. The subject has to choose the correct parts from a number of alternatives in order to make the situation logically coherent.
**Analogies**
The items consist of geometrical figures with the problem format $A:B=C:D$. The subject is required to discover the principle behind the transformation $A:B$ and apply it to figure $C$. Figure $D$ is not presented and has to be selected from four alternatives. The difficulty of the items is related to the number and the complexity of the transformations.

![Analogies example](image)

**Stories**
The subject is shown a number of cards that together form a story. The subject is given the cards in an incorrect sequence and is required to order them in a logical time sequence. The number of cards that are presented varies from four to seven.

![Stories example](image)
**Behaviour observation**

The diversity in tasks and testing materials has the advantage of making the test administration attractive for the subjects. Categories, Situations and Analogies are multiple-choice tests, the remaining four tests are so called 'action' tests. In the action tests the solution has to be sought in an active manner which makes observation of behaviour possible. Although no observation system is provided with the SON-R, many users of the SON-tests appreciate the possibilities for behaviour observation.

**Categorisation of subtests**

One can divide the SON-R into four types of tests according to their contents: abstract reasoning tests (Categories and Analogies), concrete reasoning tests (Situations and Stories), spatial tests (Mosaics and Patterns) and perceptual tests (Hidden Pictures). The abstract reasoning tests are based on relationships that are not bound by time and place; a principle of order has to be derived from the presented material and applied to new material. For nonverbal testing of abstract reasoning, classification tests and analogy tests are widely used. In the concrete reasoning tests the objective is to bring about a realistic time-space connection between objects. Emphasizing either the spatial dimension or the time dimension leads to two different test types. In the so-called completion tests (Situations), the task is to bring about an imperative simultaneous connection between objects within a spatial whole. In the other type (Stories), the object is to place different scenes of an event in the correct time sequence. The concrete reasoning tests show an affinity to tests for social intelligence in which insight in social relationships and behaviour is emphasized. In the spatial tests a relationship between parts of an abstract figure has to be established. Mosaics is a widely known test-type which was included in the earlier SON-tests; the new subtest Patterns is especially developed for the SON-R. In the perceptual test, Hidden Pictures, one must discover a certain figure hidden in an ambiguous stimulus pattern. This subtest, which is also new for the SON-tests, represents the factor 'flexibility of closure', differentiated by Thurstone.

**Memory tests**

In contrast to the earlier versions of the SON-tests, the SON-R does not include short-term memory-span tests. As Estes (1982) notes, the way information is organized and retrieved from long-term memory seems much more relevant than short-term memory in assessing the ability of children to succeed in school, where virtually all instruction is presumably intended to deal with long-term memory for the material learned.

**Characteristics of administration of the SON-R 5,5-17**

**Test administration**

Like most intelligence tests for children, the SON-R is individually administered. Group-administration is less suited for nonverbal instructions and for motivating young subjects, and would exclude behaviour observation. The role of time scoring is kept to a minimum. In this sense, the SON-R is a typical power test; there is a large variation in the difficulty of the items, while there is sufficient time for solving each item. The time needed to administer the SON-R 5,5-17 varies from 1 to 2 hours with an average of 1.5 hours. There is a shortened version of the SON-R 5,5-17, consisting of four subtests: Categories, Mosaics, Situations and Analogies. The administration of this shortened version takes about three quarters of an hour.
Verbal and non-verbal instructions

For the subtests of the SON-R there are verbal and nonverbal instructions which have been made as equivalent as possible. Nonverbal instruction forms the point of departure, verbal parts are added as accompaniment and not as supplementary information. The two sets are not intended for use as two exclusive alternatives but they give, in a different form, essentially the same information. With deaf and hearing disabled children one can often use an intermediate form by combining the nonverbal instructions with (parts of) the verbal instructions. In practice, the choice between the two procedures is generally not a problem; one adjusts to the form of communication the subject is used to.

Feedback

In two important aspects the SON-R distinguishes itself from traditional intelligence tests with regard to the test procedure: firstly, by giving feedback to the subject and, secondly, by the use of an adaptive procedure in presenting the items. It is tradition in intelligence testing not to give feedback. This tradition is broken in the SON-R because we think that such behaviour is not natural. When no reaction is allowed following an answer, the examiner's attitude can be interpreted by a subject as indifference or, erroneously, as an indication that the answer was correct. In the SON-R 5,5-17, the subject is told whether the answer was correct or incorrect following each item. However, this does not include an explanation of why an answer is incorrect. One of the advantages of giving feedback is that the subject has the opportunity to change his problem solving strategy. Also, when a subject has interpreted the instructions incorrectly, feedback offers the opportunity to adjust.

Adaptive procedure

The second important difference of the test procedure of the SON-R 5,5-17 with common test procedure concerns adaptive testing. In intelligence tests for children with a wide age range, the difficulty of the test items has to be very divergent. Presentation of all items to every subject is troublesome for a number of reasons. In the first place, this would greatly extend the duration of the test. In the second place, it is frustrating for young or less intelligent subjects to be required to solve many items that are too difficult, while the motivation of older and more intelligent subjects is reduced when they are required to solve many problems that are too easy. A practical solution, often followed, consists of presenting all items in order of difficulty and applying a discontinuation rule. However, this procedure does not result in eliminating items that are too easy for a specific subject, and the procedure has the effect that the items on which the subject fails often occur in successive order, which can be highly frustrating.

In recent years, adaptive test procedures have been developed which restrict the presentation to those items that are most suited for the specific subject. These adaptive procedures have the goal of effectively limiting the number of items to be administered with relatively little loss of reliability (Weiss, 1982). With computerized testing, these procedures can easily be implemented; with non-computerized testing, there are great practical difficulties for the examiner, both in selection and presentation of the most informative items. The SON-R uses an effective adaptive test procedure by dividing the subtests into either two or three parallel series of about 10 items. The difficulty increases relatively fast in the series. The first series of items serves to estimate the subject's general level of performance. The series is broken off after two errors. Those items in the following series that can most effectively improve and refine the measurement are administered by skipping easy items and by stopping again after two errors. This way, the administration is determined by the subject's individual performance and the presentation is limited to the most relevant items. For the examiner this method has the advantage of presenting the items within a series in a fixed sequence.
Thus, searching in the test booklet for the item which has to be presented next, takes place only at the beginning of a new series. For the subject, it is motivating that relatively easy items are presented after two errors.

**Performance of immigrant children on the SON-R 5,5-17**

In the research with hearing subjects, substantial differences in test performance on the SON-R 5,5-17 existed between immigrant children (based on country of origin of the parents) and native Dutch children. The mean IQ score for the Moroccan and Turkish children was 84, compared with a mean score of 100.5 for the native Dutch children. The lag of the other immigrant children was small (mean IQ equals 99). Comparable differences occurred in the deaf research group, with the exception of the lag of deaf children from Surinam and the Dutch Antilles which was also considerable. For deaf and hearing subjects, ethnic differences in performance concerned all subtests, but were most pronounced for Mosaics and Analogies. Neither for the hearing, nor for the deaf immigrant children did a relation exist between the number of years of residence in the Netherlands and the test scores. This indicates that lack of knowledge of the Dutch language is not an important cause of their lower results. The differences between native and immigrant children can be explained for a great part by differences in socio-economic status of the parents, as most parents of the Moroccan and Turkish migrant children belong to the lower occupational levels. The difference between native and immigrant children decreased with about one third after controlling for occupational level (see table below). Most probably the difference would decrease even more had it been possible to control for educational level of the parents as well.

**Table 6**

*Mean IQ scores per ethnic group according to occupational level*

<table>
<thead>
<tr>
<th>occupational level</th>
<th>native</th>
<th>immigrant rest</th>
<th>Morocco Turkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>17%</td>
<td>14% 89.0</td>
<td>76% 82.8</td>
</tr>
<tr>
<td>mean-low</td>
<td>38%</td>
<td>40% 100.5</td>
<td>24% 88.1</td>
</tr>
<tr>
<td>mean-high</td>
<td>21%</td>
<td>14% 96.3</td>
<td>- -</td>
</tr>
<tr>
<td>high</td>
<td>24%</td>
<td>32% 103.2</td>
<td>- -</td>
</tr>
<tr>
<td>total</td>
<td>100%</td>
<td>100% 99.1</td>
<td>100% 84.1</td>
</tr>
</tbody>
</table>
DISCUSSION

The SON-tests have been developed as an alternative to general intelligence tests for the assessment of cognitive functioning of various groups of children who are handicapped in the area of verbal communication. With the latest revisions, the SON-R tests, this has resulted in a test series which deviates from general intelligence tests in contents and in administrative procedures. In this section we will compare the SON-R both with general intelligence tests (GI-tests) and with learning potential tests (LP-tests).

The main difference between GI-tests and LP-tests is the help which is offered to the subject. In GI-tests items are presented only once, often with minimal instruction, and no training and feedback are given during test administration. With LP-tests help is given in the form of extended instructions, feedback, and training at the level at which the subject fails to succeed. The score on the LP-test reflects test performance as a result of the interactive help procedure.

Although no formal training is given in the SON-R there are several elements of the administration that facilitate learning opportunities during testing. These elements are: (a) the several examples given with each subtest, (b) the feedback which informs the subject whether the answer is correct and (c) the adaptive procedure by which easier items are presented after some failures. In this respect the SON-R shares important aspects of the testing procedure with LP-tests. The element of training is even more pronounced in the SON-test for pre-school children. In the SON 2,5-7 extensive feedback is given to the child after each failure by presenting the correct solution.

A second consideration for the comparison of tests relates to test contents and the specific abilities that are measured. Most general intelligence tests consist of a verbal and a performance scale. The verbal part, which also includes quantitative reasoning tasks, emphasizes crystallized abilities which are greatly influenced by schooling and also by more general experiences outside of school (Thorndike, Hagen & Sattler, 1986, p. 4). The performance part is more related to spatial-visualization abilities. Subtests that focus on fluid-reasoning abilities, like analogies, classification and series completion are included in either the verbal or the performance part, depending on whether the elements of the items are verbal or figural. The SON-R only contains subtests with a nonverbal content thereby excluding subtests specifically aimed at measuring verbal ability and quantitative reasoning. However, the composition of the SON-R in terms of intelligence factors is wider than the performance part of most GI-tests since it is less dominated by spatial tests. Half of the SON-R subtests are fluid-reasoning tests in a nonverbal form.

The analysis, thus far, of the different tests leads to the conclusion that the question 'SON-R, a general intelligence test or a test for learning potential?' is too simplistic; for a classification of tests more dimensions are needed. One dimension of ordering tests concerns the possibilities for learning during administration. On this dimension LP-tests score high although there is a great diversity in the amount of help and the type of training that is being offered. Traditional intelligence tests score low on this dimension and the position of the SON-R is somewhere in between. A second dimension concerns the use of a specific language in instructions and test materials. Nonverbal tests like the SON-R, LEM and the Raven aim at minimizing this aspect. A third, and very complex, dimension concerns the different cognitive aspects that are represented by the test, like verbal-, spatial- and reasoning abilities and memory, and the extent to which the measurement of these abilities depends on knowledge learned at school and/or the cultural environment. Not only between, but also within the domains of nonverbal tests, GI-tests and LP-tests, there are great variations in test composition. However, nonverbal tests are more restricted since they do not directly measure crystallized verbal ability.
The differentiation between intelligence tests is also reflected in definitions of intelligence. The aspects of knowledge, problem solving and ability to learn are stressed to different degrees, both in definitions and in tests. Which test is 'the best' can only be determined for specific situations on an empirical basis by looking at the validity with regard to relevant theoretical and practical questions. However, the comparison of tests is a very complex matter, not only between separate studies because of differences in populations and criterium measures, but also within a study it can be difficult to differentiate between the effects of reliability and the multiple factors related to content and administration on the test scores. When, for example, immigrant children score higher on test A than on test B this might be the result of differences in reliability of the tests (when standard scores are used) and not result from differences in contents and procedures.

In our opinion, the research results with the SON-R indicate that the test is a useful instrument for the nonverbal examination of children's intelligence, with high reliability and ample indications of the validity. The variety of tasks and test materials is stimulating for the subject and the adaptive procedure avoids repeated presentation of excessively difficult items. An objection to a nonverbal test like the SON-R might be that the concept of intelligence is substantially narrowed by the exclusion of verbal ability tests. However by including tests for concrete and abstract reasoning - areas that often have a verbal form in general intelligence tests - the contents of the SON-R are not limited to typical performance tests. Although the test can be administered without using language, this does not exclude the importance of verbal abilities for the evaluation of intelligence with the SON-R, as is illustrated by the correlations of the test with report marks and tests for language skills. Verbal intelligence tests often require specific knowledge learned in school. When the main object of using a test is to make predictions concerning school achievement, the absence of verbal tests in the SON-R might reduce its predictive power. If, however, the goal of intelligence assessment is to distinguish between possible causes of poor school performance, a test that is not dependent on specific knowledge is more appropriate. In such cases use of the SON-R is not only indicated for special groups such as deaf and immigrant children, but also suited for children with no specific problems in the areas of language and communication.
REFERENCES


Authors’ address:
Peter Tellegen, University of Groningen, The Netherlands, p.j.tellegen@rug.nl
Jacob A. Laros, University of Brasilia, Brazil, jalaros@unb.br