Construction and validation of the SON-R 5 1/2-17, the Snijders-Oomen non-verbal intelligence test

Laros, Jacob Arie; Tellegen, Peter Johannes

Summary

In this thesis the construction and validation of the SON-R 5½-17 is described, the recent revision of the Snijders-Oomen non-verbal intelligence test. The SON-R is an individual test of (non-verbal) intelligence for children in the ages of 5½ to 17 years. The test was published in 1989 with an extensive 'Manual and research report' in English, German and Dutch. The thesis is for the greater part a reprint of the 'Manual and research report' with the omission of norm tables and instructions for practical use of the test and with the addition of some new parts.

The SON-R is the third revision of the non-verbal intelligence test which was developed by N. Snijders-Oomen in the early forties for research with deaf children. The previous history of the test and the points of departure for the revision are described in the first chapter. The SON-R replaces two earlier revisions of the SON: the SON-'58 and the SSON (the revision by Starren in 1975). With the construction of the SON-R we aimed at combining the diversity in tasks and the possibilities of behavior observation of the SON-'58 with the psychometric qualities and standardization of procedures of the SSON. The revision of the test started in 1981; a general survey of the research is presented in section 1.3.

With all editions of the SON, the non-verbal character was of primary importance. For the administration of the test this means that neither the examiner nor the subject is required to speak or write. Applied to solving the test problems it implies that the subject needs no knowledge of a specific language. This makes the test especially suited for children with problems in the area of verbal communication and for children with little or no command of the language used by the examiner. Other common characteristics of the SON-tests concern the individual administration and the subordinate role of the time factor. The SON-tests are typical power tests with a large variation in materials and in the difficulty of the items. These common characteristics and the specific characteristics of the SON-R are described in chapter 2. Specific for the SON-R is the adaptive procedure, which adapts the administration of items to the level of the subject, and the application of feedback so that the subject knows whether the solution was correct or incorrect.

The composition of the test series is discussed in chapter 3. On the basis of their contents the following types of tests can be distinguished: abstract reasoning tests (Categories and Analogies), concrete reasoning tests (Situations and Stories), spatial tests (Mosaics and Patterns) and a perceptual test (Hidden Pictures). Three subtests are multiple choice; with the other four subtests the subject is actively engaged in constructing the solution. For several reasons memory tests were not included in the SON-R.

The construction of the seven subtests is the subject of chapter 4. The subtest Hidden Pictures holds a special position because it is the only subtest that is not composed of separate items with specific degrees of difficulty. For the construction of the other six subtests a theory of item difficulty was developed. The intention of a theory of item difficulty is to cover the most important factors which contribute to the explanation of the variation in difficulty of items in a subtest. When items are constructed according to such a theory the successive levels of item difficulty are logical subsequent steps in the mastery of the problem type specific to the subtest. On the basis of several studies, in which more than 2000 subjects participated, scales were constructed that are homogeneous to a large degree and that conform to the criteria of the Mokken model. The construction of the subtest Analogies has served as a model for the construction of the other tests. In the appendix the development of a theory of difficulty for Analogies is described and three studies are reported with a set of 60 analogy items based on this theory. The selection of the items and the conformity of the selected items to the Mokken model also receive much attention. On the basis of computer simulations an adaptive procedure was developed in which the items of a subtest are arranged in two or three parallel series. Within each series the items increase in difficulty. For each series a stop criterium is used; the starting point in the second and third series depends on the results of the preceding series. This greatly reduces the number of items administered: items that are too difficult for the subject, as well as items that are too easy, are skipped. As the variation in cognitive ability is considerable within the age-range of the test, this reduction is not only necessitated by practical considerations, but also has an important positive influence on the subject's motivation.

The standardization research is presented in chapter 5. The sample consisted of 1350 subjects, stratified according to age, sex, educational level, region, and partially according to characteristics of the

residence. For the ages of 6 to 14 years, nine groups of 150 subjects each were tested. The population has been defined as the residents of The Netherlands, residing for at least one year, and who are not suffering from severe physical or mental handicaps. Although the number of schools declining to participate was rather high, the effect on the representativeness was limited. Use of the norm tables outside The Netherlands seems mostly justified for 'western' countries with a similar cultural and socio-economic situation. Comparative research with the SON-'58 indicates a large similarity between the performance of German, Belgian and Dutch children.

The transformation of raw scores into normal distributions with a fixed mean and standard deviation requires for any age an estimate of the distribution of raw scores in the population. In order to obtain estimates for an age-group that are as accurate as possible, a statistical method was developed which simultaneously combines the information from all the age-groups. Not only is this method more accurate than the methods more commonly used (based on separate standardizations for each age-group), but it also offers better possibilities for interpolation and extrapolation outside the range of ages that have been examined. Therefore it was possible to extend the range of the test to 5½-17 years. Other relevant psychometric characteristics, such as reliability and correlations between subtests, were also fitted as a function of age in order to obtain better estimates and to make extrapolation possible (chapter 6). For the assessment of the subtest reliability a tricky problem was caused by the partial administration of the item sets as a consequence of the adaptive procedure. The usual formulas for reliability assessment could therefore not be applied. A separate study was done to examine the effect of the adaptive procedure on the reliability. The results indicated a small loss in reliability relative to complete test administration, but a considerable over-estimation of the reliability by the usual methods of reliability assessment. The improved reliability estimates of the subtests have a mean value of .76. The reliability of the total score is .93 and the generalizability of the total score (based on the relations between the subtests) is .85. The mean correlation between the subtests is .46 and increases with age. Principal component analysis indicates a strong dominance of the first factor with a large proportion of the remaining reliable variance which is specific to each subtest. Solutions with more components give support to the discrimination between spatial insight and the ability for (concrete) reasoning. The structural characteristics are largely independent of group characteristics such as hearing-deaf, native-immigrant and cognitively able-disabled.

The possibility of taking reliability and generalizability into account for the presentation and interpretation of test results is discussed in chapter 7. The common standard scores, based on standardization of the distribution of observed scores, are descriptive in nature. A special feature of the SON-R is that the distribution of the 'true' scores has also been standardized with two separate estimates of the true score for each subject. The norm score, with the corresponding standard error of measurement, is used for hypothesis testing and for group research. The latent score with the corresponding standard error of estimation, based on regression, is more suited to the evaluation of the individual ability level. Apart from these scores, which all refer to the position relative to peers, it is also possible to determine at which age the performance is 'normal'; the so called reference age. The time required for administration of the SON-R is about one and a half hours; the administration of a shortened form takes about three quarters of an hour.

Chapter 8 examines the relationship of performance on the SON-R with data gathered in the nationwide standardization research on hearing children. There were no effects of the point in time of administration on performance. Significant systematic examiner effects are in the order of 1 à 2 IQ points. Regional and local differences are small and can be explained for the greater part by differences in socio-economic background. Differences in the performance of boys and girls are slight and are absent for the total score. However, socio-cultural factors are of importance: mean IQ-scores are clearly related to parental occupational level and to parents' native country. The lag in IQ-scores of the immigrant children appears to be related to their lower socio-economic background. Extensive research was done on the relations of test performance with educational variables. The multiple correlation of three aspects of the school career - type of school, grade repetition and report marks - with the IQ scores is .59. The correlation is .66 with a school achievement test at the end of primary education. The validity of the test, related to performance at school, is quite satisfactory. Certainly, if one wants to make a distinction between intelligence and school achievement, in which intelligence is one of the important factors that influence future school achievement, the absence in the SON-R of verbal tests that highly depend on learned skills and school knowledge can be judged favorably.

The first SON-test was developed for research with deaf children. The aptness of the test for this specific group has had high priority in the subsequent revisions. The SON-R was administered to almost the entire population of deaf children in the Netherlands from 6 to 14 years. The results are presented in chapter 9. As with hearing subjects

there are no important differences between boys and girls, while differences related to socio-cultural background are important and comparable in magnitude. The mean IQ score of the deaf children is considerably lower compared to the hearing children; a difference which remains after correction for the high proportion of immigrant children with the deaf. Subsequent analysis made it seem likely that these differences do not primarily arise as a result of the auditory handicap but as a result of cognitive handicaps which are found next to deafness in about a quarter of the deaf population. For three quarters of the deaf population there are only slight differences with the hearing population; these differences are mainly related to the tests for abstract reasoning. Amongst others, this was a reason not to standardize IQ scores separately for the deaf, although the IQ score can be converted to a percentile score for this population. The correlation of the SON-R with earlier administrations of SON-tests is .76. The fact that the IQ scores according to the new norms are considerably lower is of great importance for use of the test in practice. This is not only a result of the lack of separate norms for the deaf, but is for the greater part the result of the general finding that old test norms are less adequate as the level of cognitive performance improves in the population over the years. For the deaf, there appears to be a strong relation of the SON-R with verbal ability; the correlation of the SON-R with a written language test is .49. Within the deaf population there is a minor relation of test results with hearing loss while the relation with onset of deafness and cause of deafness is clear. The results indicate that a non-verbal intelligence test like the SON-R makes it possible to do useful and valid research on the cognitive functioning of deaf children.

The instructions for administration of the test (chapter 10) are limited in this thesis to the inclusion of general instructions, descriptions of the items, and to the complete instructions for one subtest. In the general instructions the adaptive procedure and the way feedback is given are described in detail. Each subtest has both verbal and non-verbal instructions that are printed next to each other. The non-verbal instructions are points of departure; it is not the purpose of the text to give additional information.

The last chapter, chapter 11, is a practical guide for the use of the scoring form, the norm tables and the computer program that is included with the test. The adaptive procedure and points specific to each subtest are printed on the scoring form, as a memory aid. Because of the adaptive procedure, scoring is done during the administration of the test; the answer keys are indicated on the form. There are two alternatives for the computation of the standardized scores. Either one uses the norm tables which are included in the 'Manual and research report' for 38 different age groups, or one uses the computer program which computes all results given the age and the raw subtest scores and subsequently prints these and saves them on file. An important advantage of the computer program is the saving of time and the elimination of computing and writing errors. Also the results will be more accurate as they are based on the exact age and because some computations can be performed by hand only in a simple and statistically less sophisticated way. A third advantage is that the possibilities of the program for the computation of scores are more extensive, especially the possibility to compute standardized total scores for any combination of subtest scores. Finally, in chapter 11, some concrete examples on the use and interpretation of the scores of the SON-R are presented.