

University of Groningen

## De ontwikkeling van een nieuwe versie van de S.O.N. voor 7-17 jarigen

Starren, Johannes Christianus Maria Gerardus Marcellus

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

1973

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Starren, J. C. M. G. M. (1973). *De ontwikkeling van een nieuwe versie van de S.O.N. voor 7-17 jarigen*. s.n.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## SUMMARY ')

The research of which an account has been given in the chapters above is typically a piece of applied scientific work. It is not primarily aiming at an exploration of a field of research or at the testing of hypotheses, but at the construction and gauging of an instrument of research, a series of intelligence tests. This objective determines the course of the whole work and the nature and place of each partial examination which has been carried out. It is always a matter of selecting and testing experimentally the materials and forms of research. It is true that more than once a suggestion or a starting point has been found into the direction of questions more theoretical and, in themselves, important, but deliberately these have not been worked out further.

The test-series which is to result from the research should meet several special requirements apart from the general test-technical qualifications like: objectivity, normization, reliability and validity. Pre-eminently it must be suited to the examination of early-deaf people; for that reason subtests were developed which do not require the use of language for instruction and reaction, with which individuals can be examined and in which no limits of time are handled. Since individual, non-verbal tests are also important for the examination of hearing persons, a verbal parallel instruction has been made and used in the gauging of people whose auditory faculty is normal. On a more theoretical level the new test-series is to render possibilities for generalization to a comprehension of intelligence which is as extensive as possible; the exclusive use of performance tests does not fit into this plan, because these practically only examine perception and analysis of spatial forms.

The whole set of requirements for the intelligence test-series, to be constructed anew, until this moment differs but very little from the one used for the S.O.N.-versions of 1943 and 1958. The reasons why we decided to develop a new version of the S.O.N. (short: S.S.O.N. 7-17) are not so much lying in differences of starting-points but in the elaboration of them. First of all there are doubts concerning the partial construct validity of the intelligence concepts, 'form', 'concrete coherence', 'abstraction' and 'memory' from the precious version: it is true that the intercorrelations between the subtests, from the concepts that are different for a priori reasons, are far from perfect, but that is not saying enough, especially when we take into account the relatively

' ) Translated by Drs. J. Willems.

low reliability coefficients of some subtests. These data, placed against the background of objections which have often been expressed and are more practical, like the relatively long duration of an examination and the difficult manageability of the material belonging to the examination, have led to the question whether the elaboration cannot be more satisfactory. Apart from this we are confronted with the situation that the previous version has been in use for more than ten years now since it was gauged, with all the risks of a certain obsolescence: probably there are changes in the subpopulations for whom the test-series was intended, items become outdated and are therefore less recognizable now, etc.

The course of the research project as a whole can be sketched in the following outlines. The general form that was given to the test-series (chapter 2, par. 2.1.), was determined from studies of literature and experience with the version that has been in use till now. On account of this the following conditions were formulated, which the subtests that are to be reconstructed, respectively to be developed, must meet as much as possible. The test-series must enable people to make an objective, that is to say, reproducible, judgement concerning the maximum persons can perform. For that reason we have tried to minimize the role of the test-leader as 'interpreter' and observer - by means of a logical and empirically verified score-system and a procedure of examination which is as uniform as possible. By replacing all loose objects that were packed in large and small stationary cases by items in a couple of ring-files we expect to shorten the duration of the test. The test-series must be built up out of homogeneous subtests; this means that in the process of selecting the items special attention must be paid to the degree in which the items measure the same thing. Another measure meant to heighten the reliability of subtests, concerns the insertion of a large number of similar and attractive items. There must be the possibility that the test-series can be applied individually and without time-limits. Care must be taken - in order to guarantee that individual differences resulting from these series can be attributed, as many as possible, to differences in 'intelligence' - that people need not have a specific knowledge of or experience with the material used to find the right solution. For this reason we have abandoned words and numbers as possible contents of items; or else particularly early-deaf persons will be placed at a disadvantage right from the start as compared with people who can hear.

In the choice of the subtests (chapter 2, par. 2.2.) one can rely on the literature and the experiences with the previous S.O.N.-versions to a far less extent. I have not explicitly sought for a linking-up with recent theories concerning

intelligence, because these bear upon the structure and the functioning of 'intelligence' of people who can hear and, in so far as an arbitrary selection would have been made, operationalization in an instrument that is also useful in practice does not seem to be possible at short notice. The S.O.N.-versions from 1943 and 1958 supply certain starting-points indeed, but on numerous points research is necessary to be able to arrive at decisions that can be accounted for. For the present we consider the a priori concepts 'form', 'concrete coherence', 'abstraction' and 'memory' with Snijders and Snijders-Oomen, as useful labels to make clearer what one must imagine a concept like 'intelligence' to be. We still have our doubts because an empirical verification of the psychological meanings of these concepts has all but failed to come.

In the first stage of the research we concentrate our attention mainly on improvement of reliability and shortening of the duration of examination of the existing operationalizations. A series of preliminary investigations with revised and also some newly constructed subtests have been carried out in order to check, more specifically, the effect of changes in the ways of offering and amount of items; in order to gain an insight into the level of differentiation and difficulty of the items, in the mutual coherence of the subtests etc. In this way a tentative battery of eight subtests comes into being that can be considered for use. Compared with the previous S.O.N.-versions a profit has been gained as to duration of examination and reliability. By means of structure analysis of the inter-correlations of the subtests some indications have been found for the partial validity of the chosen labels of 'intelligence'

In the main investigation before the gauging these eight subtests were applied to a sample of 237 persons: 197 seven-year-old and fourteen-year-old who could hear and forty seven - and fourteen - year-old early-deaf persons, in the course of which care was being taken to have sufficient spreading as regards degree of urbanization, sex, nature of education and social-economic level. On the level of items and subtests reliability indexes were calculated. The intention is, on the one hand, to consider whether the amount of subtests must be curtailed further and on the other hand to improve the subtests themselves, and if possible to shorten them. On account of the results one of the subtests was omitted and changes were introduced to improve further the build-up as to difficulty and the homogeneity of the remaining subtests. The results give us sufficient grip to suppose that the test-series particularly meets the requirements of reliability and efficiency. Less convincing are the validity results. In order not to rouse possibly false suggestions beforehand

concerning the psychological meaning of the seven subtests a name-giving was chosen that connects as directly as possible with the things the child has to do - in the order of examination: (1) Composing a figure, (2) Making a strip, (3) Finding analogies, (4) Remembering pictures, (5) Mirroring a figure, (6) Completing situations, (7) Classifying pictures.

Raw scores have hardly any significance, although they are the result of reliable and efficient measuring. The gauging-stage (Chapter 3) of a research offers the condition for a better granting of significance to test scores: (a) by comparing individual scores with accurately defined normgroups (Chapter 3, par. 3.1.) and (b) by transformation of raw scores into standard scales (Chapter 3, par. 3.2.).

On the ground of indications from literature, which were confirmed later by our investigations, we expect that age as well as hearing as principles of division must be used for a standardization. In view of the greatness of the differences between the average raw scores of succeeding ages in the samples for gauging, it is not justified to carry out the gauging of age for shorter periods than one year. The selection of the hearing persons takes place according to the model of a three-plane, stratified sample with geographical strata in the first level (sample of municipalities) and in the lower levels, stratification according to nature of education (sample of schools) and age and sex (sample of persons). The names of the Dutch municipalities have first been grouped according to degree of urbanization and geographical situation. From the list thus obtained twenty five municipalities were chosen and that with a probability, which is proportionate to the number of inhabitants. From the files of various authorities eighty persons were selected from each municipality, equally divided over the two sexes and the ten age-groups and in such a way that a division, determined beforehand, concerning the kinds of education or training would be realized. During the drawing from the files we also paid attention to the relation between date of birth and test-date: only the persons whose dates of birth are at most one month before or after the test-date were considered. During the gauging on early deaf people there is no question of drawing samples: we confine ourselves to all 7-17-year-old people from the five Dutch institutions for the deaf; with the exception of those who cannot be considered because of illness or on account of the nature of their handicap (e.g. autism). The original plan to compose a sample for early-deaf persons which would be maximally comparable to such a sample for hearing people, on reflection meets with considerable objections of practical and methodical nature.

In September 1970 the examination of the S.S.O.N. 7-17 started: on the one side on a sample of 2011 hearing Dutch children and

on the other hand on 754 pupils of the Dutch institutions for the deaf. By the middle of 1971 we brought this to a conclusion; an enterprise to which more than eighty test-leaders lent their assistance. Neither the amount of refusals to undergo the test examination nor the deviations of samples with respect to the distribution of population are such and to such an extent that we can speak of an essential impairment of our original plan.

In transforming raw scores into some system of standard scores, such great measures of spreading are often imposed on the new score-distributions that the effect of the average reliability of the tests is not taken into account sufficiently. We therefore have in mind to use, among others, a school marking-scale, a scale that better combines the demands for differentiation, exactness and possibility to interpret the result than e.g. the traditional IQ and Wechsler-scales. The transformation to these scales was done after the model of the standard distribution for the significant, not normal distributions; for the raw score-distributions that cannot be considered to be not-normal the standard scores were calculated via linear transformation. To give the user of the test a freedom of movement that is as great as possible, total scores were worked into norm-tables for various subtests combinations.

In the last section of Chapter 3 (3.3.) an account is given of the elaboration of the gauging data. The central question in this respect is whether the items, the system of items in subtests, the structure of the test-series, the relation to personal, environmental and educational variables justify the conclusion that the new version of the S.O.N. is a test that offers reliable and valide information, in the shortest possible time, with respect to some aspects of 'intelligence' with early deaf and hearing 7-17-year old persons.

The results of the item-analyses prove that it is possible to construct subtests for a relatively extensive range of age with a satisfactory internal consistency; it is especially for the total scores that relatively high (internal consistency) reliability coefficients are found. Neither the subtest-total correlations nor the results from factoranalytic research lend support to the theoretical structure of the test-series: the operationalizations of the four a priori intelligence concepts appear to be mutually interchangeable. For all age-groups, deaf and hearing persons we do best to confine ourselves to a description of the observed variance in one dimension. To make plausible that the S.S.O.N.-total scores measure 'intelligence' we checked, by comparisons between groups and by correlation calculation, whether the test scores discriminate, respectively cohere with external variables to such an extent as is expected. So the gauging

samples on the ground of variables of person and environment (age, hearing, sex, municipal type, geographical situation, occupation of the father and previous schooling of the parents) have been divided into groups of which has been assumed that they will differ, respectively will hardly or not differ in 'intelligence'. The investigation of correlations of S.S.O.N.-scores with variables of schooling (type of school, fluency, marks at school, judgement of the school) gives a more direct contribution to the determination of the construct validity of the S.S.O.N.: to these variables themselves validity for the construct 'intelligence' can be ascribed. Besides judgements of social behaviour at school have been drawn into the investigation of correlations, for the very reason to answer the question how far the S.S.O.N. measures other constructs than 'intelligence', which is naturally not our intention. Generally speaking the results of the comparative and correlational research mentioned above are in conformity our expectations. So we are inclined to infer that the S.S.O.N -scores possess validity with respect to 'intelligence'.

We may be brief about the last chapters. Chapter 5 starts with a detailed description of the tests, succeeded by the most important quantitative characteristics of each subtest. In Chapter 6 we discuss the practice of examining and scoring. The book concludes with a critical survey of the possibilities and particularly the impossibilities of the test-scores, so that the revised S.O.N. will be used in a most justifiable way (Chapter 7).