Bottom-up rehabilitation in schizophrenia
Appelo, Martinus Theodoor

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:
1996

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

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).

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.

Download date: 31-01-2020
5. Selection of instruments for rehabilitation research.

5.1 Introduction.

The aim of this chapter is to operationalize the theoretical starting points, as described in the first two chapters, with the help of (neuro)psychological and psychiatric instruments and tests. The initial choice of variables will be motivated by explaining their relevance in the process of learning and training. Also, the variables for evaluation of the present study will be selected, based on research into the reliability of selected instruments in a sample of schizophrenic patients. The need for such a selection is twofold.

Firstly, related to the theoretical starting points, the variables will range from basic cognitive functions to demographic variables. This implies that many variables will be included. However, a precondition for statistical evaluation in a relatively small sample, is a rather limited amount of variables. According to Bartko et al. (1988) the soundest data reduction technique is prevention of data proliferation in the design phase by selecting a few crucial measures with established reliability.

The second reason is mainly concerned with cognitive variables. Regarding the fact that most cognitive disturbances are characterized by strong fluctuations, Deelman (1990) maintains that in today's neuropsychology the issue of test-retest reliability is strongly underestimated. Therefore, he recommends a test-retest coefficient of .70 as the absolute minimum for individual diagnostic research and argues that high test-retest reliability is necessary in all clinical research. With regard to reliability-research in schizophrenia, it can be concluded that significant improvement in multiple areas of neuropsychological functioning occurs during the recovery period following acute episodes of the illness (Sweeney et al., 1991). This finding strengthens the need for stable variables for evaluation research.

It may be argued that using a control group, as will be done in the present study, reduces the need for stable variables for evaluation of the study. However, results of both experimental and control groups can be influenced by 'spontaneous' fluctuations, which can only be eliminated by using very large samples, which will not be the case in the present study.

The first domain of variables is concerned with cognitive functioning. Based on Anderson's model (1983) variables will be selected for operationalization of the declarative stage, knowledge compilation and production tuning. The second domain involves stress and coping. As stated before, at the behavioural level, stress and coping are closely related since the quality of coping is supposedly related to the subjective evaluation of the stressor (Lazarus & Folkman, 1984). Furthermore, the number of stressors experienced was repeatedly found to be independent of the level of stress reported (Reich et al., 1988; Brantley et al., 1988). Therefore, stress will not be operationalized separately, but coping-styles will.

Functioning at the behavioural level will be operationalized further by selecting variables for competence of social skills performance and general behaviour as observed by others. As far as schizophrenic symptoms are concerned, positive and negative symptoms will be operationalized. Finally, some demographic variables will be selected to...
evaluate the practical implications of the present study.

5.2 Operationalization of theoretical starting points.

5.2.1 Cognitive functions.

Central to the declarative stage is the working memory. Cognitive functions central to this stage and described in section 2.2., are attention, perception and memory. A precondition for learning new skills or relearning old skills is an optimal use of these functions. Tests selected to measure these components of information processing are the Continuous-Performance-Test (CPT; Rosvold et al., 1956), the Stroop-Color-Word-Test (Hammes, 1971), GIT-Picture-Completion (Luteyn & van der Ploeg, 1983), WAIS-Picture-Arrangement (Wechsler, 1981), WAIS-Digit-Span (Wechsler, 1981), Rey-Verbal-Memory-Test (Rey, 1964) and GIT-Word-Fluency (Luteyn & van der Ploeg, 1983).

Information processing: Continuous Performance Test (CPT).

The Continuous Performance Test was developed to examine the ability to sustain focused attention over time within a rapidly paced visual discrimination task. This task is generally used in the cognitive study of schizophrenia.

Several studies found performance deficits on this task in schizophrenic patients (Nuechterlein & Dawson, 1984b). CPT-results are relatively stable and to some degree unrelated to psychiatric symptoms. Therefore some authors assumed that CPT-results represent a stable vulnerability indicator of schizophrenia (Nuechterlein et al., 1990). Also a positive correlation between CPT-results and long term course of schizophrenia is assumed (Cornblatt & Erlenmeyer-Kimling, 1984; Erlenmeyer-Kimling & Cornblatt, 1987). However, not all schizophrenic patients perform poorly on this task; about 40-50% of the patients perform poorer than the poorest performance of normal controls (Orzack & Kormentsky, 1971). In the poor-performance group, the incidence of having a relative with schizophrenia is higher as compared with the schizophrenic patients who perform better, and the course of schizophrenia is unfavourable in this group (Buchsbaum et al., 1990). Given this state of affairs, patients with poor performance on the CPT, are presumably profiting less from rehabilitation efforts considering that impaired vigilance will disrupt the learning process despite the use of tactics as described in chapter 3.

The computerized CPT-version used in this study (Nuechterlein et al., 1986) is an adaptation of the original CPT. The task is to detect and respond to a target stimulus sequence (a 7 only when it follows a 3) by pressing a button. In this study 600 numbers are presented with 1000 msec interval and 80 msec stimulus presentation. The target stimulus combination (3-7), is presented 90 times. The performance measure is the sensitivity index (CPT-d’), which summarizes the subject's discrimination of targets from nontargets and is closely related to the level of arousal (Davies & Parasuraman, 1982).

Nuechterlein et al. (1990) investigated the stability of this 'memory load' version
of the CPT over a one year period. Seventeen schizophrenic patients were tested in clinical remission and during an active psychotic state. Significant deficits in CPT-d' were detected in both clinical states. However, the degree of deficiency clearly became greater during the psychotic state. Test-retest coefficient was low ($r=0.33$).

Van den Bosch et al. (submitted) measured test-retest coefficients of this version in a sample of 21 schizophrenic patients, 16 major depressive patients, 19 nonpsychotic nondepressive psychiatric patients and 20 normal controls. The test was administered three times. The interval between the first and second assessment was one week, and between the second and third assessment 11 weeks. All coefficients were high (first-second assessment: .89; second-third assessment: .86; first-third assessment: .85). In their study the authors also discuss the validity of this CPT-version. CPT-d' score did not differ significantly between schizophrenic and depressive patients and high correlations were found between reaction time variability and CPT-d'. They conclude that the sensitivity measure of a double-stimulus CPT measures response variability rather than the maintenance of a high level of attention during a prolonged period of time.

Selectivity of attention: Stroop-Color-Word-Test (STROOP).

The STROOP measures selectivity of attention; the ability to focus on one source of information and to actively inhibit response to information of other sources. An inability to inhibit irrelevant information during training has important consequences for training and its material surroundings.

The STROOP includes three cards. On the first card 100 colournames (yellow, green, red and blue) are printed in 10 straight lines. Subjects have to read the card as quickly as possible. On the second card 100 coloured blocks are printed. Subjects have to name the colours. The third card is like the first. However, on this card words are printed in different colours (e.g. 'red' is printed in yellow). Subjects have to name the ink colours in which the color words are printed, ignoring the linguistic information that does not match the ink colour. Time needed for the second card is subtracted from the time needed for the third card. This score, which will be used in this study, is called 'interference score' (STROOP-INT) and involves interference from the well learned semantic associations to the printed words that intrude automatically (Nuechterlein & Dawson, 1984b). When selectivity of attention decreases, this score will increase.

Abramczyk et al. (1983) and Carter et al. (1993) found that schizophrenic patients performed poorly on this test. Asarnow et al. (1977) found no significant performance impairment among offspring of schizophrenic mothers on the interference condition of this test. Brink et al. (1984) found a test-retest coefficient of .69 in a normal sample ($n=99$), with a test-retest interval of two months. Van der Schoot et al. (1986) found a significant higher score on STROOP-INT at re-assessment compared to the first assessment in a sample of 53 normal subjects.

Gestalt-formation: GIT-Visual-Closure.
Accurate visual perception of information is a prerequisite to learning skills, especially when modelling or any other form of visual instruction is used as a training strategy.

The GIT-Visual-Closure-Test measures the 'speed-of-closure' or gestalt-formation. Twenty successive drawings of objects, animals and persons, are presented in an ascending order of difficulty. Subjects have to name what the particular drawing represents. Performance is rated by counting the correct responses (GIT-VC).

Luteyn and van der Ploeg (1983) found a high split-half-reliability on this task in a normal sample (n=1570, r=.84). Lackroy (1966) calculated the 'delayed' split-half reliability in a sample of 40 students; Second half of the test was administered ten days after the first half. A reliability coefficient of .77 was found. Test-retest reliability was calculated by Lackroy (1966) in a sample of 43 students (interval 4 weeks) and Leppers (1981) in a sample of 363 students having learning-difficulties and 306 normal controls (interval 3 years) respectively. Results show coefficients of .86, .80 and .82. respectively.

Social perception: WAIS-Picture-Arrangement.

Perception, organization, categorizing and arrangement of interpersonal and emotional information are important functions in every social situation. For example during skills training social perception, among other factors, contributes to the degree in which the trainer is understood.

WAIS-Picture-Arrangement consists of ten series of pictures that have to be rearranged to form a logical story. The ten series have an ascending order of difficulty and an increasing number of pictures. Subjects have to perceive, judge and interpret social interactions on the pictures to be able to compose the right story. Performance is scored by counting the correct responses (WAIS-PA).

According to Penn (1991) the Picture-Arrangement test has the best face validity of commonly used neuropsychological tasks for assessing nonverbal social perception. Patients with frontal dysfunctions tend to non-response behaviour on this task (Lezak, 1983). Only global, first impressions and a rough analysis of the pictures, are used to compose the story.

Reliability of this WAIS-subtest has been calculated repeatedly in a psychiatric inpatient sample including schizophrenic patients. Results show split-half reliabilities of .78 (n=50; Ryan et al., 1982), .84 (n=75; Boone, 1991) and .83 (n=100; Boone, 1992). Test-retest reliability in a sample of 21 male psychiatric and neurological patients was calculated by Ryan et al. (1985). Results show a test-retest reliability of .79 with a mean interval of 38 weeks. However, results at the second assessment were significantly higher compared to first assessment. Improvement at the second assessment was also found by Moore and Stambrook (1990) who found a test-retest reliability of .69 in a sample of patients with head-injuries (n=60, interval 8.5 months).

Working memory capacity: WAIS-Digit-Span.
Learning, remembering and reproduction of declarative information is central to the first stage of skills acquisition. Dysfunctions in the working memory capacity disrupt the learning process and may require special mental prosthetics.

WAIS-Digit-Span is focused on the working memory capacity operationalized by reproduction of declarative information. The first part of this test consists twelve series of digits. The number of digits ascends from 3 to 8. Numbers are presented one by one, with an interval of one second. Subjects have to represent the series. In the second part, the digit span ascends from 2 to 7 and subjects have to represent the series in reverse. Total score is the sum of correctly represented series on both trials (DIGITS).

Reliability of this WAIS-subtest was calculated twice in a psychiatric inpatient sample by Boone (1991; 1992). In both studies he found a test-retest coefficient of .81 in a sample of 30 and 25 patients respectively with an interval of 2-3 weeks and 2-8 weeks respectively. Moore and Stambrook (1990) found low test-retest stability on this test in a sample of patients with head-injuries (n=60). Stability coefficient was .61 and results improved at test-retest interval (34 weeks). Ryan et al. (1985) found a test-retest reliability of .75 and no improvement in a sample of 21 male psychiatric and neurological patients (interval 38 weeks). A study with chronic schizophrenic patients (n=13, interval 28 weeks) was administered by Prescott et al. (1986). They found a high test-retest coefficient (r=.79) and no differences in performance. Exactly the same results were found by Sweeney et al. (1991) in a sample of 39 schizophrenics, with an interval of one year.

Verbal memory: Rey Verbal Memory Test.

The Dutch version of Rey's auditory verbal learning test (Deelman et al., 1980) measures short and long-term verbal memory functions.

The test consists of 15 mono syllabic nouns, which are presented five times on audiotape, with an interval of one second. After each presentation, subjects have to reproduce the words they remember. The score (RALT-C) is the added total of correct recalls, and reflects the ability to learn the words. About 15 minutes after the last presentation, filled with a non-verbal task, subjects have to recall the words they remember (RALT-R). This score expresses the coding in long-term memory and the ability to retrieve the information. Finally a recognition task is administered in which 30 words, including the 15 'original' words, are presented. Subjects have to decide whether or not a particular word was present in the original list (RALT-RE).

Differences in recall and recognition scores are believed to have strong implications for training. In schizophrenia, as described in section 2.2., recognition memory is often intact whereas recall memory is deficient.

In a normal sample (n=106) test-retest reliability of both copy and recall is high; .80 and .83 respectively (interval 2 months; Heslinga, et al., 1983a). However, Van der Schoot et al. (1986) found both scores to improve significantly between assessment and reassessment.

Prescott et al. (1986) administered a word recall and recognition task, using 16 words and two trials for the recall-task and a list of 48 words for the recognition task, in a sample of 14 chronic schizophrenics. Interval between two assessments was 28 weeks. Test-
retest coefficients were .57 for recall and .67 for recognition. No differences in scores between assessment and re-assessment were found. Sweeney et al. (1991) calculated performance stability and test-retest reliability coefficients of RALT-C and RALT-R in a sample of 39 schizophrenics (interval one year). Test-retest coefficient for RALT-C was .63, while no significant change was found. RALT-RE had a coefficient of .48, while significant improvement between first and second assessment was found.

Search in long term memory: GIT-Word-Fluency.

Searching and retrieving information from long term memory is essential for reproducing already existing and newly learned production systems. The GIT-Word-Fluency test is related to this. The test consists of two parts; first the subject has to name as many animals as possible in one minute. In the second part they have to name as many professions as possible in the same time. Both parts have separate scores, GIT-WF1 and GIT-WF2 respectively.

Luteyn and van der Ploeg (1983) found a split-half-reliability of .76 on this task (sum of both scores) in a normal sample (n=1570). Lackroy (1966) calculated the 'delayed' split-half reliability in a sample of 40 students. A reliability coefficient of .62 was found for the sum of both scores. Test-retest reliability was also calculated by Lackroy (1966) in a sample of 43 students (interval 4 weeks). Results show coefficients of .68 (animal-naming) and .59 (profession-naming). Schouten et al. (1978) found a test-retest reliability of .75 (animal-naming) and .53 (profession naming) in a normal sample (n=39), without significant changes on both scores between first and second assessment (interval 3-4 months).

Central to the second stage of the learning process is proceduralization and composition. One task was selected to measure this component of information processing; The 'Tower of Hanoi' (Simon, 1975).

Proceduralization: Tower of Hanoi.

Learning how to solve a problem implies learning an effective procedure for steps to follow. If one effectively learns how to solve a problem by trial and error, a shorter procedure can be established the next time one is faced with the same problem.

The Tower of Hanoi is a 'puzzle' consisting of three poles mounted on a rectangular board. Disks of varying diameter and colour are slipped on the left pole with the largest disk at the bottom and the smallest at the top. Subjects are instructed to move the tower from the left to the right pole as quickly as possible in as few moves as possible. They are not allowed to move more than one disk at a time nor to place a larger disk on a smaller one. Following Schmand (1991), who found that psychiatric patients were not able to solve a five disk puzzle, first a three disk trial (HANOI-3) and immediately thereafter a four disk trial (HANOI-4) is presented. About 20 minutes later, filled with a verbal task, subjects have to repeat this test. Moves to find and complete the required procedure are scored. If subjects do not reach a solution, a preset limit is scored: 40 moves in the three disk trial and
80 moves in the four disk trial.

No reliability results are available of this test. Schmand (1991) applied this test in a sample of 80 psychotic patients and 27 normal controls. The number of steps required to complete the 3 and 4 disk condition did not diminish significantly in the second presentation within the same assessment session.

The Tower of Hanoi is a 'test-retest' in itself. To prevent the effect of testing, the scoring-procedure is adjusted for the present study. HANOI-3 and HANOI-4 are respectively the sum of moves needed to complete both 3 disk trials and 4 disk trials within the same assessment session.

Using feedback is prerequisite for generalization and discrimination at the stage of production tuning. In order to measure the adequate use of external feedback, the Wisconsin Card Sorting Test (Grant & Berg, 1948) was selected:

Using feedback: Wisconsin Card Sorting Test.

Rehabilitation implies skills-training. Trainers have to correct ineffectively executed production-systems by using feedback. The learning-process will be disrupted if patients are unable to profit from corrective feedback.

The Wisconsin Card Sorting Test (WCST) is a task which measures concept-formation as a function of learning by feedback. In this study the test is presented by computer. The subject has to sort cards with printed figures according to colour, form or number of figures. The test includes four target cards: one red triangle, two green stars, three yellow crosses and four blue balls. A fifth card is shown which matches with three of the four target cards in one feature. This card has to be sorted without instruction on which dimension. Feedback ('right' or 'wrong') is given after every single sort. After ten correct sorts, the sorting principle is changed. The subject has to sort 64 cards and sorting principles are fixed (colour, form, number, colour, form, number). One score is administered: WCST-perseverations (WCST-P), the sum of errors that immediately follow an identical error, despite feedback. This score will represent the degree of adequate use of feedback, since this will lead to few perseveration errors.

In general, patients with schizophrenia perform poorly on this test (Goldberg et al., 1987; Goldberg & Weinberger, 1988;). Pogue-Geile (1990) and Franke et al. (1992) found a higher frequency of perseverations in a sample of healthy siblings of schizophrenic probands as compared to healthy controls. Consequently Franke et al. (1992) argue that WCST-perseverations is a vulnerability marker to schizophrenia. However, in their study WCST-perseverations were unrelated to the severity or duration of the illness.

According to Seidman et al. (1991) no attention had been paid to the question of stability of performance on this test until they concluded their own study. They found a test-retest coefficient for WCST-perseverations of .39 and no significant differences in a sample of 12 schizophrenic patients (mean interval 3 years) and a coefficient of .87, and no differences in a sample of 11 patients (mean interval 6 months). The low correlation in their first study is related to within-subject variation on the test. This finding signifies that some patients are not able to learn from feedback.

Sweeney et al. (1991) calculated performance stability and test-retest reliability
coefficients of WCST-P in a sample of 39 schizophrenics (interval one year). Test-retest coefficient was .73, while significant improvement between the first and second assessment was found.

5.2.2 Coping-style.

Psychiatric patients in general are deficient in using effective coping styles (Van den Bosch et al., 1992). In order to achieve the goals of rehabilitation as formulated in section 1.4., it is necessary to develop control on stressful situations. To initiate appropriate behavioural strategies when faced with a problem, a problem-oriented coping-style should be present. Therefore, emotion-oriented coping, such as avoidance and other fear reducing measures related to the judgment of not having control, should be diminished while problem-oriented coping should be trained.

One Dutch test operationalizing coping-styles is available, and will be used in this study; the 'Utrechtse-Coping-Lijst' (UCL; Schreurs et al., 1988).

Coping-styles: Utrechtse-Coping-Lijst.

The UCL is a questionnaire containing 47 items. Each item represents a specific way of dealing with a problem. Subjects are asked how often they apply the particular strategy when faced with stress or problems in general. Response on each item is scored on a 4-point Likert-type scale: hardly ever (1), sometimes (2), often (3) or very often (4). In a normal sample 7 factors were identified, each representing a different coping style (Schreurs et al., 1988). One factor unambiguously represents problem-oriented coping: Problem-Solving (UCL-P). This factor includes seven items such as; 'solving the problem step by step', 'applying goal-directed behaviour' and, 'thinking it over before acting'. Two factors are clear operationalizations of emotion-oriented coping: Avoidance (UCL-A), including eight items like 'waiting for things to happen' and 'avoiding the problem', and Depressive-Reactions (UCL-D), including seven items such as 'constantly worrying' and 'having the feeling nothing can be done about the problem'. These three factors have been selected for the present study.

In the present study the UCL will be presented as an interview; to prevent possible misunderstandings leading to invalid responses, the items are read to the patient. In case of doubt about the patients' comprehension of a particular item, it is explained by the interviewer.

Regarding intercorrelations between the factors, Schreurs et al. (1988) found a coefficient of -.13 and -.10 for 'Problem-solving - Avoidance' and 'Problem-solving - Depressive-Reactions' respectively. 'Avoidance - Depressive-Reactions' shows an intercorrelation of .37. Internal consistency coefficients in a normal sample (n=1200) are moderate to high (Problem-solving .82, Avoidance .73, Depressive-Reactions .70). Test-Retest reliability coefficients, calculated in a sample of 84 students (interval 6 weeks) are .79, .64 and .69 for Problem-Solving, Avoidance and Depressive-Reactions respectively.
5.2.3 Social skills.

Poor social functioning, as stated before, is a characteristic symptom of schizophrenia. As in most treatment programmes, improving social competence by means of social skills training will be a goal in the present study.

In order to assess the social competence of psychiatric patients, Bellack et al. (1990) strongly recommend the use of role-play tests. According to the authors, who found strong support for the concurrent validity and reliability of a role play test in a psychiatric sample, these tests are inexpensive, easy to implement and applicable to the vast majority of patients. Moreover, molar ratings of competence are easy to administer and account for a large proportion of variance in performance (e.g. Merluzzi & Biever, 1987).

One of the best validated behavioural tests for the measurement of social competence is the Simulated Social Interaction Test (SSIT; Curran, 1982). This test was selected for the present study.


The SSIT, adjusted by Mersch et al. (1992) for use in a Dutch population, consists of eight short simulated interactions measuring responses to: disapproval or criticism, social assertiveness or visibility, confrontation and anger expression, heterosexual contact, interpersonal warmth, conflict with or rejection by parent or relative, interpersonal loss, and receiving compliments.

In each situation, a narrator describes the social situation by means of audio-tape and a confederate delivers a prompt in vivo. The subject has to react to this prompt. Performance of the patients is recorded. Video-tapes are rated by trained judges, in this study clinical psychology students.

Competence of performance (SSIT-C) is rated qualitatively at a global level using 11-point Likert-type scales ranging from 'no or completely irrelevant reactions' (1) to 'very assertive' (11). Assertiveness is defined as: standing up for one's opinion and taking into account the other one's feelings and opinion (score 9, 10, 11). If a subject only stands for his or her opinion without offending the other, '7' or '8' is scored. A '6' is scored if the subject tends to stand up for his opinion and does not offend the other. A non-assertive reaction (score '4' and '5') can be sub-assertive; only taking into account the other one's feelings and opinion without standing for one's own, or aggressive (score '2' and '3'); a response offending the other and obstructing meaningful continuation of communication. Since eight situations are played, competence-score ranges from 0 to 88.

Besides competence, asking W-questions (SSIT-W) as a reaction to the prompts, is rated since this will be explicitly trained in the social skills training which is part of the rehabilitation programme. In four out of eight situations, reacting with a W-question, is rated as assertive (9-11).

Several studies support the interrater reliability and internal consistency of the performance score. Interrater reliabilities across studies reported by Mersch et al. (1992) are invariably high (.90 or higher) for trained raters. In their own study they found an
interrater coefficient of .91 in a sample of 74 social phobic patients.

The use of trained judges allows for a high degree of control over what is observed. This preference might lead to problems because of the possible social irrelevance of the judges' ratings. Related to this issue Wallander et al. (1983) used different rater groups, including a sample of untrained community representatives and an untrained group of research assistants to evaluate social performance on the SSIT by normals and psychiatric patients. Results show high levels of similarity in judgments.

Curran (1982) calculated test-retest reliability of the SSIT, rated by two trained judges, in a sample of 33 psychiatric patients (interval 6 months). Test-retest coefficients of both raters were high for the competence score (respectively .88 and .91). No differences were found between the first and second assessment.

5.2.4 General functioning.

A low level of role-functioning is one of the diagnostic criteria for schizophrenia. Of course, improving general functioning is a goal of rehabilitation efforts. As to reviewing instruments that assess the functional living skills of chronically mentally ill patients, Wallace (1986) concludes that none of them are entirely adequate. However, according to the author, one of the most carefully developed and reliable instruments is the 'Rehabilitation Evaluation Hall and Baker' (REHAB; Baker & Hall, 1983), which will be used in the present study because it is also easy to assess.

General behaviour: REHAB.

REHAB, adjusted by Van der Gaag and Wilken (1994) for use in a Dutch psychiatric population, was designed to assess the institutionalized patient's deviant and general behaviour. It consists of 23 items that are completed by nurses who base their ratings on behaviours and events that occurred during the week before the evaluation. In the present study only the 16 'general behaviour' items will be used. They measure speech characteristics, social interaction, engagement in leisure time activities, activity level, table manners, personal hygiene, grooming, care of personal possessions, caring for own needs, managing money, using public facilities, and overall functioning. Each item is rated by making a mark along a line which is scored afterwards using a transparent rating-sheet with 10-point Likert-type scales ranging from 'the worst possible performance' (9) to 'appropriate performance for independent living' (0). Maximum score on the general behaviour subscale is 144, signifying the worst level of general behaviour.

According to Wallace (1986) thorough psychometric investigation has resulted in good reliability and validity of REHAB. Baker and Hall (1988a; 1988b) found that the interrater reliability of each item ranged from .61 to .92 (n=47 psychiatric patients). Also the general-behaviour score is found to be sensitive to change and differentiates between day hospital patients and long-stay inpatients. In two other studies moderate to acceptable interrater reliability of items was found; Conway (1989) found coefficients ranging from .65 to .73 (n=307) and Croucher et al. (1989) found a range of .55 to .78 (n=625). The same
holds for the Dutch version of REHAB. Van der Gaag and Wilken (1994) found coefficients ranging from .58 to .82 in a sample of 598 patients, each rated by two psychiatric nurses. The test-retest coefficient of the general-behaviour score (n=583) was .96. Also internal consistency, measured with Cronbach's alpha was high (.89).

5.2.5 Symptoms.

In section 1.1. the distinction between positive and negative symptoms of schizophrenia was described. The former are characteristic of psychotic episodes, the latter are predominant in between these episodes.

Positive symptoms will be measured with the Schizophrenia Subscale of the Comprehensive Psychopathological Rating Scale (CPRS; Asberg et al., 1978). Negative symptoms are measured with the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1983).

Positive symptoms: CPRS-SS.

The CPRS is a clinical interview consisting of 67 items, of which 40 are rated on the basis of the patient's answers during the interview, while the remaining 27 items concern the examiners' observations of the patient's behaviour during the interview. Rating is based on behaviours and symptoms that occurred during the month before the interview. According to Asberg et al. (1978), the items may be grouped into subscales according to the needs of a particular study.

In the present study the Schizophrenia Subscale will be used (Luckner & Woggon, 1985; Poelijoe et al., 1987). The 17 items of this subscale include: derealisation and depersonalisation, formal thought disorder, delusions (of control, of reference, paranoid and other delusions), hallucinations (hallucinatory behaviour, commenting voices, auditory and other hallucinations), inadequate affect, black outs, muscle tensions, diminished understanding of speech, incoherent speech and mannerisms. Each item is rated using a 4-point Likert-type scale ranging from 'absence of the symptom' (0) to 'extreme occurrence of the symptom' (3). Range of the total-score (CPRS-SS) is therefore 0-51.

Inter-rater reliability at item level was calculated repeatedly. Kuny et al. (1982) found Kappa's ranging from .27 ('other delusions') to 1.00 ('black-outs') in a sample of 30 psychiatric patients. The Dutch version of the CPRS was investigated by Goekoop et al. (1991) in a sample of 99 psychiatric patients. Satisfactory Kappa's (≥ .61) were found for 8 of the 17 CPRS-SS items. Reasonable Kappa's (.41 - .60) were found for 2 items and moderate to low Kappa's (≤ .40) for 4 items. 'Other hallucinations', 'incoherent speech' and 'hallucinatory behaviour' appeared insufficiently present in this study to calculate Kappa's.

Negative symptoms: SANS.
According to Andreasen (1989) the SANS was developed in order to facilitate the evaluation of five important symptoms frequently observed in schizophrenia. These symptoms include affective flattening, alogia or poverty of speech and thought, avolition, anhedonia, and attentional impairment. Each of the five symptoms is broken down into observable behavioural components which are rated on a 6-point Likert scale ranging from 'absence of the symptom' (0) to 'extreme occurrence of the symptom' (5). Since the SANS includes 30 items, range of the total-score (SANS) is 0-150.

The SANS is rated during a clinical interview but information is based on multiple sources of information; direct observations by the interviewer and nurses, and reports from relatives and the patient himself. Rating is based on behaviours and symptoms occurring in the month before the interview.

The inter rater reliability for the global ratings of the five subscales is consistently high in a variety of cultural settings. Andreasen (1989) reports Kappa's ranging from .63 to .95, and a mean inter rater reliability coefficient of .84 for the total score (Andreasen, 1982). It is stressed that high reliability cannot be achieved without adequate training of the interviewers.

5.2.6 Demographics.

An important aim of the present study is the prevention of long-term hospitalization of schizophrenic patients with a duration of illness longer than 5 years. All subjects were initially referred to a long stay ward, but were offered the possibility to participate in the rehabilitation programme.

Concerning the relation between demographic variables and course of schizophrenia it is a known fact that the amount of former admissions predicts re-admissions (Harding et al., 1987a,b; Slooff, 1988). Some follow-up studies report that women have better prognosis than men (Harding et al., 1987a,b; McGlashan, 1988).

Concerning correlations between demographic variables and cognitive functions, van der Gaag (1992) concluded that performance on cognitive tests (including CPT, Picture Arrangement, Stroop, WCST, Word Fluency, and Visual Closure) covaries with age and level of education, but not with duration of illness and dose of antipsychotic medication. These results were also reported by Schmand (1991).

In the present study, age, sex, duration of illness, former admissions and level of education will be included as independent variables. Age and duration of illness will be rated in years, whereby the latter is calculated from the first time a psychotic episode appeared. Educational level will be rated using a Dutch system (Heslinga et al., 1983b); a 7-point scale ranging from unfinished elementary school (1) to university (7).

Since antipsychotic medication has little effect on cognitive measures and generally normalizes the patient's cognitive functions, it will not be controlled in the present study (van der Gaag, 1992). According to Schmand (1991) there is no simple and elegant solution for controlling drug effects. To study drug free patients might seem the best solution, but in the present study, which only includes chronic patients, this is only a theoretical matter. Therefore, concerning medication a practical stand will be taken. Patients participating in the programme will be optimally treated before the programme.
starts. Patients needing a major change in antipsychotic medication will be excluded from the study and benzodiazepines and anticholinergics may only be changed temporarily within a small range of dosage. This practical stand is also defensible because in clinical practice and in everyday life of schizophrenic patients outside the hospital, this strategy is more or less a constant factor in episode-free periods.

In order to evaluate the practical effect of the study, the place of referral following treatment is included, as well as the 'living-situation' at six months and one year follow-up. At the follow-up the 'working-situation' will also be assessed.

5.3 Selection of variables.

As stated in section 5.1 in today's neuropsychology the issue of test-retest reliability is strongly underestimated. Deelman (1990) recommends a test-retest coefficient of .70 and argues that high test-retest reliability is necessary in all clinical research. However, in the case where mean change occurs in test scores, the reliability coefficient does not reflect the stability of scores in absolute terms, but the degree in which individuals maintain their relative position in the distribution of scores or, the extent to which the degree of change in test performance is consistent across the sample (Sweeney et al., 1991). Therefore, the primary statistical analyses necessary to assess performance stability should be testing the hypotheses that mean scores do not differ and calculating test-retest reliability coefficients. Only tests with high test-retest coefficients (≥ .70) and without significant differences between the first and second assessment can be used in evaluation research.

With regard to cognitive variables as described in the previous section, only WAIS-Digit-Span does meet the criteria for evaluation-research in a psychiatric sample. This variable shows unambiguously high test-retest coefficients (.79 to .81) and in most studies no improvement at the second assessment was found in a sample of psychiatric inpatients. Only three other variables have been investigated in a psychiatric sample on both test-retest coefficients and changes in mean scores. Rey-Verbal-Memory-Test shows low to moderate coefficients (.48 - .67) on all variables, and contradictory results have been reported as to changes in mean scores. The same has been found for WCST-perseverations, although coefficients range from low to high (.39 - .87). WAIS-Picture-Arrangement shows moderate to high reliability coefficients in different samples including psychiatric patients (.69 - .84), but improvement at the second assessment has frequently been reported. Concerning CPT-d' only contradictory results have been reported (test-retest coefficients .33 and ≥ .85). Data pertaining to the Tower of Hanoi only show that no differences can be found between mean scores. With regard to the other variables (STROOP-INT, GIT-Visual-Closure and GIT-Word-Fluency) no reliability-results in a psychiatric sample could be found.

Concerning non-cognitive variables, no data are known to be available of the test-retest reliability of the UCL and SSIT in a schizophrenic sample. The REHAB, CPRS-SS and SANS have repeatedly proven to be reliable at item-level in a psychiatric sample, including schizophrenic patients. Reliability results of these variables concern inter-rater reliabilities, since test-retest results are always obscured by the fact that these measures
concern a span of time before the moment of assessment. Considering that in the present study the total score of CPRS-SS, SANS and REHAB (general behaviour) will be used, these results are not sufficient to include these variables without further research.

In general, results reported here do not permit the use of these variables for evaluation of the present study. Therefore, a study was administered in a sample of schizophrenic patients, to select variables meeting the criteria (1) a test-retest reliability coefficient of $\geq .70$ and, if so (2) no differences in test and retest mean scores. With regard to CPRS-SS, SANS, and REHAB an inter-rater reliability of $\geq .70$ is the single criterion.

To obtain a representative sample, patients were selected and asked to participate at two different settings at different locations in the Netherlands; One setting for young patients (Academisch Medisch Centrum, Amsterdam), admitted after first or only few episodes of psychosis, and one long-stay ward for patients with chronic schizophrenia (Psychiatrisch Ziekenhuis, Franeker). All patients met DSM-III-R (APA, 1987) criteria for schizophrenia. Patients were paid for their participation (10 Guilders/hour).

Patients were tested twice by trained clinical psychology students, supervised by a clinical psychologist. To obtain inter-rater reliability of CPRS-SS and SANS, all patients were interviewed by a clinical psychologist and a psychiatrist, working in the same hospital and trained together in conducting the interviews. Inter-rater reliability of the REHAB, is based on observations by untrained nurses.

The SSIT-performance, recorded on video, was rated by three trained clinical psychology students. As the inter-rater reliability of the three rater-pairs at the first assessment was very high (.97, .98, .99 for SSIT-C and 1.0 for SSIT-W) the test-retest results were based on the mean scores of the three raters.

Initially the total sample consisted of 27 patients. All patients were closely observed by ward-personnel during the test-retest interval of two weeks. In case of marked changes in a patient's clinical picture, that particular patient did not participate at re-assessment. Two patients showed changes in psychotic symptoms, one patient did not complete all tests at both assessment sessions. As a result, a sample of 24 patients completed both the assessment and re-assessment sessions. In table 5.1 the demographics of this sample are described.

Test and retest data from all variables except CPRS-SS, SANS and REHAB were correlated (Pearsons correlations) and if a coefficient of $\geq .70$ was found, means were tested with Paired-Samples-T-Tests. Inter-rater results of CPRS-SS, SANS and REHAB were correlated (Pearsons correlations).

Results, presented in table 5.2, 5.3, and 5.4 show that according to the pre-set criteria CPT-d', DIGITS, HANOI-3, WCST-P, UCL-P, SSIT-C and SSIT-W, CPRS-SS, SANS and REHAB are reliable. These results warrant that these variables can be used in schizophrenia research aimed at evaluation of rehabilitation efforts.

The rehabilitation programme and the control conditions, that will be evaluated with these variables, are described in detail in the next chapter.
Table 5.1: Demographics of the test-retest sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean (s.d.)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>16</td>
<td>27.1 (9.1)</td>
<td>18-49</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
<td>4.5 (1.3)</td>
<td>1-7</td>
</tr>
<tr>
<td>Age in years</td>
<td>24</td>
<td>7.3 (8.3)</td>
<td>1-34</td>
</tr>
<tr>
<td>Level of education</td>
<td>24</td>
<td>3.2 (3.9)</td>
<td>0-15</td>
</tr>
</tbody>
</table>

Table 5.2: Selection of variables by correlations (n=24).

<table>
<thead>
<tr>
<th>Variable</th>
<th>T1</th>
<th>T2</th>
<th>Correlation (Pearson)</th>
<th>Two-Tailed Probability</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT-d'</td>
<td>2.40</td>
<td>2.65</td>
<td>.798</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>Stroop-int</td>
<td>51.5 (1.3)</td>
<td>43.7 (29.0)</td>
<td>.589</td>
<td>.004</td>
<td>-</td>
</tr>
<tr>
<td>GIT-VC</td>
<td>13.2 (2.6)</td>
<td>14.4 (14.4)</td>
<td>.886</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>WAIS-PA</td>
<td>11.2 (5.0)</td>
<td>13.0 (4.9)</td>
<td>.750</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>DIGITS</td>
<td>11.6 (4.8)</td>
<td>11.4 (4.1)</td>
<td>.875</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>RALT-C</td>
<td>37.1 (11.3)</td>
<td>46.3 (14.5)</td>
<td>.818</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>RALT-R</td>
<td>6.8 (3.1)</td>
<td>9.5 (3.3)</td>
<td>.850</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>RALT-RE</td>
<td>28.3 (2.5)</td>
<td>29.0 (2.3)</td>
<td>.565</td>
<td>.005</td>
<td>-</td>
</tr>
<tr>
<td>GIT-WF1</td>
<td>17.4 (16.0)</td>
<td>13.3 (3.7)</td>
<td>-.13</td>
<td>.247</td>
<td>-</td>
</tr>
<tr>
<td>GIT-WF2</td>
<td>19.7 (6.5)</td>
<td>17.5 (5.3)</td>
<td>.551</td>
<td>.005</td>
<td>-</td>
</tr>
<tr>
<td>HANOI-3</td>
<td>23.6 (15.3)</td>
<td>23.0 (15.1)</td>
<td>.803</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>HANOI-4</td>
<td>73.2 (37.8)</td>
<td>79.7 (49.5)</td>
<td>.310</td>
<td>.141</td>
<td>+</td>
</tr>
<tr>
<td>WCST-P</td>
<td>11.6 (13.5)</td>
<td>10.8 (13.7)</td>
<td>.927</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>UCL-P</td>
<td>16.2 (4.4)</td>
<td>16.0 (4.3)</td>
<td>.853</td>
<td>.000</td>
<td>+</td>
</tr>
<tr>
<td>UCL-A</td>
<td>19.2 (4.6)</td>
<td>18.5 (3.2)</td>
<td>.609</td>
<td>.003</td>
<td>+</td>
</tr>
<tr>
<td>UCL-D</td>
<td>15.1 (4.1)</td>
<td>15.3 (3.6)</td>
<td>.661</td>
<td>.001</td>
<td>+</td>
</tr>
<tr>
<td>SSIT-C</td>
<td>48.4 (3.8)</td>
<td>47.1 (4.7)</td>
<td>.843</td>
<td>.001</td>
<td>+</td>
</tr>
<tr>
<td>SSIT-W</td>
<td>84 (-77)</td>
<td>68 (-82)</td>
<td>.754</td>
<td>.001</td>
<td>+</td>
</tr>
</tbody>
</table>
Table 5.3: Selection of variables by testing means (n=24).

<table>
<thead>
<tr>
<th>Variable</th>
<th>T-value</th>
<th>Two-Tailed Probability</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT-d'</td>
<td>-1.51</td>
<td>.145</td>
<td>+</td>
</tr>
<tr>
<td>GIT-VC</td>
<td>-4.30</td>
<td>.000</td>
<td>-</td>
</tr>
<tr>
<td>WAIS-PA</td>
<td>-2.57</td>
<td>.017</td>
<td>-</td>
</tr>
<tr>
<td>DIGITS</td>
<td>0.44</td>
<td>.662</td>
<td>+</td>
</tr>
<tr>
<td>RALT-C</td>
<td>-5.38</td>
<td>.000</td>
<td>-</td>
</tr>
<tr>
<td>RALT-R</td>
<td>-7.59</td>
<td>.000</td>
<td>-</td>
</tr>
<tr>
<td>HANOI-3</td>
<td>0.32</td>
<td>.751</td>
<td>+</td>
</tr>
<tr>
<td>WCST-P</td>
<td>0.77</td>
<td>.448</td>
<td>+</td>
</tr>
<tr>
<td>UCL-P</td>
<td>0.36</td>
<td>.724</td>
<td>+</td>
</tr>
<tr>
<td>SSIT-C</td>
<td>1.00</td>
<td>.329</td>
<td>+</td>
</tr>
<tr>
<td>SSIT-W</td>
<td>0.83</td>
<td>.420</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 5.4: Inter-rater Reliability (n=24).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rater1 mean (sd)</th>
<th>Rater2 mean (sd)</th>
<th>Correlation (Pearson)</th>
<th>Two-Tailed Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>REHAB</td>
<td>39.9 (18.2)</td>
<td>37.4 (15.4)</td>
<td>.748</td>
<td>.001</td>
</tr>
<tr>
<td>CPRS-SS</td>
<td>11.0 (5.7)</td>
<td>11.4 (5.5)</td>
<td>.971</td>
<td>.000</td>
</tr>
<tr>
<td>SANS</td>
<td>63.1 (24.5)</td>
<td>61.5 (25.5)</td>
<td>.953</td>
<td>.000</td>
</tr>
</tbody>
</table>