Order out of chaos?
Assessment and treatment of executive disorders in brain-injured patients
Order out of chaos?
Assessment and treatment of executive disorders in brain-injured patients

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen in het openbaar te verdedigen op dinsdag 28 juni 2011
om 13.30 uur precies

door

Daniëlle Henriette Elisabeth Boelen
Geboren op 24 april 1975
te Maastricht

© D.H.E. Boelen 2011
All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means
without permission in writing of the authors, or, when appropriate, from the publishers of the publications.
The research described in this thesis was carried out at the St. Maartenskliniek Nijmegen, in collaboration with the Radboud University Nijmegen, Donders Institute for Brain, Cognition, and Behaviour.

This study was supported by the Dutch Organisation for Health Research and Development (ZonMw), Rehabilitation Research Programme (Grant No. 1435.0009).
Contents

Chapter 1 General introduction 9
Chapter 2 Executive dysfunction in chronic brain-injured patients: Assessment in outpatient rehabilitation 23
Chapter 3 Script generation in brain-injured patients with a dysexecutive syndrome 49
Chapter 4 Rehabilitation of executive disorders after brain injury: Are interventions effective? 69
Chapter 5 Effects of a multifaceted treatment programme for executive dysfunction after acquired brain injury on indications of executive functioning in daily life 127
Chapter 6 Predictors of treatment outcome in a successful experimental treatment of executive dysfunction 155
Chapter 7 General discussion and conclusions 167
Chapter 8 Samenvatting 181
Dankwoord 187
Curriculum Vitae 191
Donders Graduate School for Cognitive Neuroscience Series 193
General introduction
“Human cognition is forward-looking, proactive rather than reactive. It is driven by hopes, plans, goals, ambitions, and dreams, all of which pertain to the future and not to the past. These cognitive powers depend on the frontal lobes and evolve with them. The frontal lobes endow the organism with the ability to create neural models as a prerequisite for making things happen, models of something that, as of yet, does not exist but which you want to bring into existence.” (Goldberg & Bougakov, 2007, p. 345).

Executive functions: a historical perspective

Early physiological investigations have shown that stimulation of the prefrontal cortex does not produce motor or sensory reactions, and that damage to this region does not produce paralysis or paresis, nor does it produce impairments of sensation or speech. Therefore, investigators have long regarded the prefrontal cortex as a ‘silent’ area. However, animal studies showed that extirpation of the frontal lobes led to apathetic, passive and inhibited behaviour (see Luria, 1969). Investigations in humans led to the recognition of two main symptoms: a disturbance of complex forms of active, purposive behaviour, and a disturbance of the critical attitude towards the patients’ own defects (Luria, 1969). It was concluded from such studies that the frontal lobes play an important role in the organization or programming of behaviour.

Luria (1966) postulated three functional units in the brain, of which the first is a subcortical unit, dedicated to the regulation of a ‘mental tone’ and a state of consciousness. It corresponds to the cognitive concept of arousal, which is defined as a general state of receptivity and preparedness to respond (van Zomeren & Brouwer, 1994). Secondly, Luria distinguished the posterior part of the cerebral cortex that is responsible for the processing of sensory information. The third unit consists of the frontal lobes, of which the prefrontal regions are the most anterior portion and which are held responsible for the programming, regulation and verification of goal-directed behaviour. Luria proposed that the prefrontal areas of the brain are superimposed on all other cortical areas, enabling the prefrontal regions to perform a more global and supervisory function of regulation of behaviour. Or, in the words of Luria’s pupil Goldberg: “The frontal lobes are to the brain what a conductor is to an orchestra, they coordinate and lead neural structures in concerted action.” (Goldberg, 2001, p.2). Luria (1966) described that damage to the frontal lobes resulted in deficits in problem-solving, decision-making, and active thinking. These symptoms were part of the so-called frontal lobe...
syndrome, a disorder concerning the overall organisation of cognition and action. Although considerable variations exist in the patterns of symptoms observed in patients, two main components of the frontal lobe syndrome were recognized. The first includes a ‘disturbance of incentive’, comprising a narrowing of interests and a disorder of initiative, while the second consists of a general disinhibition, a disturbance of emotion and character, and a gross defect of critical attitudes.

Cognitive theories postulating a central control system: unitary theories

Luria’s ideas about the frontal lobes already pointed toward their role as a central executive mechanism in the regulation of behaviour. More recent theories have stressed this role of the prefrontal regions in executive control. Norman and Shallice (1986), for example, have developed an information processing model that sheds light on this executive character of the prefrontal lobes. Two main assumptions underlie the model. The first is that the control of selection of routine actions and thought operations is decentralized, which is called ‘contention scheduling’. Thus, within routine situations the appropriate response is selected by automatically activating schemata based on contextual information. However, in non-routine situations such automatic responses are not available or insufficient, so that a higher control mechanism is necessary. In Norman and Shallice’s model, this control mechanism is called the Supervisory Attentional System (SAS), which modulates the information processing and response selection in new, non-routine situations. The system is ‘dysexecutive’ when the response selection in non-routine situations is impaired.

In performing many complex tasks it is necessary to hold information ‘online’ in temporary storage to complete the task. Baddeley (1986, 2000) has proposed a theory in which the role of working memory within executive functioning is clarified. The model involves an attentional controller (the ‘central executive’), aided by three subsidiary slave systems. The visuo-spatial sketchpad holds and manipulates visual images, the phonological (or articulatory) loop performs a similar function for speech-based information and the episodic buffer links information across domains to form integrated episodic representations of visual, spatial, and verbal information with chronological ordering, such as the recollection of a story or event. The episodic buffer is also assumed to have links to long-term memory and semantical meaning. The central executive is a flexible system responsible for the control and regulation of cognitive processes. It can be thought of as a supervisory system that facilitates or inhibits cognitive processes and thus is very similar in function to the SAS. Several functions have been attributed to the central executive, such as the coordination of the slave systems, selective attention and inhibition, binding information from various sources into coherent episodes, and shifting between tasks (i.e. multitasking).

Duncan (1986) has developed a theory on the control of behaviour by its desired results, collating the ideas and observations of Luria (1966) and Bianchi (1922). Duncan proposes that any activity requires that a list of goals be used to generate a structure of action by which those goals are achieved. In solving problems only those actions are selected that reduce the difference between current states and goal states. A plan of action consists of a hierarchical structure of successive sub-goals that unfolds as a result of the requirements and constraints on behaviour. Duncan suggests that the symptoms of the frontal lobe syndrome represent a defect in the process of reducing the mismatch between current states and goal states. Thus, disorganisation of behaviour is the consequence of a failure in the ability of goals to seize and maintain control of the course of actions until they have been achieved. The behaviour of individuals suffering from a frontal lobe syndrome may be characterized by an inability to retain a stable purpose of behaviour on the one hand and the insertion of irrelevant actions on the other hand. Such irrelevant or ill-judges intrusions or the apparent neglect of task demands is called goal-neglect (Duncan, Emslie, Williams, Johnson, & Freer, 1996) and leaves behaviour fragmented, volatile, passive, incongruous and superficial.

Fractionation of executive control: multiple process theories

In contrast to the earlier unitary views, recent approaches have emphasized the multifaceted nature of executive functions and attempts have been made at dividing these functions into various putative subcomponents (Baddeley, 2000; Miyake, Friedman, Emerson, Witzki, & Howeter, 2000; Shallice, 2002; Stuss, 2006; Stuss et al., 2002). Clinical practice has shown that dissociations exist between tasks assessing various executive (sub)functions, e.g. a particular patient may exhibit deficits in planning and organising, but not in inhibition, while another patient may show problems with initiative and cognitive flexibility, but not with planning skills. Also, research has shown that different frontal regions may be responsible for distinct processes,
Another recent idea, mainly based on imaging and lesion research studies, is that different executive functions might not only be mediated exclusively by the frontal areas, but depend on other, non-frontal brain regions as well. The findings indicate that the brain has an executive system that is hierarchical and yet distributed rather than strictly localized. Lesion research suggests that dysexecutive behaviour is not always the general result of frontal lobe damage. In many cases patients who show symptoms of the ‘frontal lobe syndrome’ have damage outside the (pre)frontal cortex, that sometimes includes other cortical and/or subcortical structures (see for example Elliot, 2003; Godefroy, 2003; Heyder, Suchan, & Daum, 2004; Jurado & Rosselli, 2007). The prefrontal cortex is connected to the basal ganglia and the cerebellum via the thalamus in cortico-subcortical circuits, which are held to serve as neuroanatomical substrates of executive processing. Elliot (2003) stresses the roles of the basal ganglia and striatum in executive functioning. Executive deficits are seen, for example, in patients with Huntington’s disease and Parkinson’s disease as a result of pathology in these subcortical regions (Elliot, 2003). Redgrave, Presscott and Gurney (1999) propose that the basal ganglia provide the brain with a central selection mechanism specialized to resolve conflicts over access to limited motor and cognitive resources. The thalamus plays an important role as a relay, which is connected with multiple brain regions, such as the prefrontal cortex. Patients with thalamic lesions may show executive deficits, e.g. problems with planning and inhibition (van der Werf et al., 2003). Clinical reports have also described impaired executive functioning after damage to the cerebellum (Schmahmann & Sherman, 1998; Schweizer et al., 2008). Anatomical and neuroimaging studies have identified the likely basis for these impairments in a cortico-ponto-cerebellar network through which the cerebellum is connected to the anterior areas of the brain (Bellebaum & Daum, 2007). Moreover, functional neuroimaging studies have shown that parietal areas play a critical role during the performance of executive tasks and that executive processes such as shifting, updating and inhibiting are not subserved by the frontal lobes alone (for a review see Collette, Hogge, Salmon, & van der Linden, 2006). Another region of the brain that has been associated with executive functions is the anterior cingulate cortex (ACC), on the medial surface of the frontal lobe. Functional neuroimaging data suggest that the ACC serves a strategic function in the selection of action, by reducing the competition between potential responses to a stimulus. Another view concerning the role of the ACC comes from electrophysiological studies of the error-related negativity (ERN). ERN-studies have suggested that the ACC is part of a circuit involved in error detection and compensation. Subsequent
functional neuroimaging data support the idea that the ACC detects conflict between incompatible response tendencies. Thus, the evaluative function of the ACC detects processing conflicts and as such is part of an error prevention network (Carter et al., 2000; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). In short, not only has a multi-component view of EF been developed, but considering the prefrontal cortex as the sole substrate of executive processing is progressively replaced by the idea that a more distributed circuitry of cortical and subcortical brain regions mediates successful executive functioning.

The dysexecutive syndrome: a clinical perspective

An attempt to circumvent the aforementioned localization problem was the introduction of the concept of ‘dysexecutive syndrome’ by Baddeley (1986). This was a functional definition of executive deficits, rather than a classification of patients on the basis of an anatomical localization. The dysexecutive syndrome consists of numerous cognitive, emotional and behavioural symptoms. Although many of these symptoms regularly co-occur, it is common to encounter patients who have several, but not all of these symptoms. Impairments in executive functioning hamper the ability to successfully manage the problems of everyday living. Patients are unable to appreciate daily life situations, to recognise personal strengths and weaknesses and to identify opportunities and threats. The deficient appreciation of situations hinders the ability to reflect, to evaluate and judge a reasonable course of action, anticipating the possible outcomes of decisions, and finally, to review this outcome and make further adjustments and plans as necessary. An example of a clinical description of executive sub-functions comes from Ylvisaker (1998). He distinguishes eight aspects of executive functions: self-awareness of strengths and needs, realistic and concrete goal-setting, planning the steps to these goals, self-initiating these plans, self-monitoring and evaluating performance according to plan and goal, self-inhibiting behaviour not leading to the goals set, flexibility and problem solving when situations cannot be dealt with according to plan, and finally, strategic behaviour, i.e. transfer of successful behaviours to other situations. In brain-injured patients these aspects can be differentially impaired, leading to different patterns of dysexecutive symptoms. The above mentioned aspects do not include social awareness and social adaptation. Nevertheless, executive functions also play a crucial role in complex social behaviour (e.g. Barkley, 2001; Stuss & Levine, 2002).

Outline of this thesis

The functional definition of executive dysfunction is clinically relevant, because it emphasises the impairments rather than stressing the lesion localisation. Such a definition implies that numerous patients with various aetiologies and lesions in different brain regions may exhibit symptoms of executive dysfunction. These quantities outnumber those based on frontal lesions alone. Executive dysfunction can result from head trauma, cerebro-vascular disease, cerebral tumours, cerebral infections such as encephalitis, hypoxia, and degenerative diseases. Executive disorders are also found in a range of psychiatric conditions including schizophrenia, attention deficit disorder and personality disorders. The executive complaints of a wide range of brain-injured patients have probably as yet been ill-understood and underestimated, so that many of their problems may not have been assessed and treated adequately. In rehabilitation clinicians regularly encounter patients and their proxies who complain about executive problems in daily life. Many of these patients do not have lesions in frontal brain regions. The aim of this thesis is to assess and treat executive problems in this large population of brain-injured patients. Two main research questions will be addressed: first, can the complaints of these patients be evidenced by neuropsychological tests, questionnaires and observation scales, and second, can the daily life executive problems of this particular patient population be treated effectively?

Chapter 2 presents an exploratory study on the assessment of executive disorders in a sample brain-injured patients referred for outpatient rehabilitation. It addresses the question whether the complaints of patients and/or their proxies about executive problems in daily life, are tapped by neuropsychological tests, questionnaires and observation scales. The study examines the sensitivity of several tests and questionnaires to verify the experienced executive problems. These instruments are used to assess executive dysfunction in patients who later participated in the effect study described in chapters 5 and 6.

Chapter 3 presents a study that investigates script processing in brain-injured patients with executive disorders. Script generation abilities of patients are compared to those of healthy control subjects. Script generation serves as a measure of planning skills. Performance on various types of scripts (i.e. open-ended vs. closed and low vs. high frequency scripts) are evaluated within these two groups of subjects.
Chapter 4 summarizes the results of a critical literature review of existing treatments of executive disorders. The objective of this review is to assess the internal and external validity of studies that have investigated the effect of specific interventions for executive impairments. Thus, this review aims to scrutinize the methods used in various efficacy studies and the verification of their corresponding results.

Chapter 5 describes a prospective randomized controlled trial of the efficacy of a new, multifaceted treatment program for patients in outpatient rehabilitation with executive disorders. This multifaceted treatment relied heavily on implications for treatment that follow from the abovementioned theoretical frameworks of executive functioning proposed by Luria (1966, 1969), Norman and Shallice (1986), Baddeley (1986, 2000), and Duncan (1986). This experimental treatment is administered to a group of patients diagnosed with executive disorders. The effects of the experimental treatment are compared to those of a matched group of patients with executive disorders who received a standard cognitive treatment. Both groups are assessed before, immediately after, and six months after treatment.

Chapter 6 provides additional results about our efficacy study described in chapter 5. Specifically, the question is raised whether and which pre-treatment characteristics of the experimental group can predict their successful outcome after treatment.

Chapter 7 is dedicated to a joint discussion of the results of the studies described in this thesis. Several methodological issues of the studies are addressed and both theoretical and clinical implications are suggested.

References

Transactions: Biological Sciences, 351, 1405-1412.
Executive dysfunction in chronic brain-injured patients: Assessment in outpatient rehabilitation

Published in:
Neuropsychological Rehabilitation, 2009; 19, 625-644
D.H.E. Boelen, J.M. Spikman, A.C.M. Rietveld, L. Fasotti
Abstract

In this study 81 chronic brain-injured patients referred for outpatient rehabilitation, who complained of executive impairments in daily life situations and were observed by proxies and therapists to have such problems, were assessed using various tests and questionnaires of executive functioning, such as the BADS and the DEX Questionnaire. The main purpose was to examine the sensitivity of these instruments in this particular group of patients. The tests and the DEX were also administered to healthy controls to investigate which of the instruments discriminate optimally between patients and healthy controls. The results indicate that the tests as well as the questionnaires were sensitive to the executive problems of the patients. There were no significant differences between DEX ratings of patients, proxies and therapists. This suggests that patients who were eligible for outpatient rehabilitation showed relative intact awareness into their executive problems. A specific combination of three ‘open-ended’ tests and the DEX contributed significantly to the prediction of group membership.

Introduction

Many brain-injured patients referred for outpatient rehabilitation have difficulties with planning, problem-solving and reasoning. These difficulties can be characterized as executive deficits, which can vary from relatively mild to rather severe. Although the fact that outpatient rehabilitation referral implies that patients are, at least partially, able to resume their previous activities, even subtle executive deficits may profoundly affect the individual’s ability to return to the community setting (Crepeau & Scherzer, 1993; Vilkki et al., 1994). Therefore, for this particular group, outpatient rehabilitation is often aimed at stimulating further social and vocational adjustment. Despite the fact that executive problems are frequently seen in this rehabilitation population, relatively little is known about the severity of these deficits and the best way to measure them.

One way to investigate the presence of symptoms is by means of questionnaires and rating scales. With these methods, self-observations by patients and observations made by others, such as relatives and carers, can be obtained in a structured and standardized way, in order to provide information about the patients’ behavioural and emotional changes. Questionnaires such as the Dysexecutive Questionnaire (DEX; Burgess, Wilson, Evans, & Emslie, 1996) are designed to sample a broad range of daily problems commonly associated with executive disorders. Besides executive questionnaires, general rating scales and interview formats of competency in daily pursuits and social behaviour can be used to evaluate dysexecutive behaviour (for example, the Patient Competency Rating Scale; Prigatano et al., 1986). However, it must be kept in mind that such measures are, by definition, subjective, being based on judgements of overall frequency of occurrence of problems rather than on specific counts of well defined behaviours. These reports can be inaccurate, because patients tend to underestimate their problems (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Prigatano & Schacter, 1991; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). For this reason, observations by significant others, such as partners or professionals, are essential to assess executive impairments more accurately (Bennett, Ong, & Ponsford, 2005a; Burgess et al., 1998; Knight, Alderman, & Burgess, 2002; Norris & Tate, 2000; Wilson et al., 1996). Discrepancies between patients’ reports and significant others’ reports are considered as a clinical marker of the patients’ awareness of executive dys functioning (Burgess et al., 1998).

In order to obtain more objective indications of executive problems, a wide variety of neuropsychological tests is available, such as the Tower of London Test (Shallice, 1982), verbal fluency tests (for example, Benton &
Hamsher, 1976), the Twenty Questions Test (Laine & Butters, 1982), and the Stroop Test (Stroop, 1935). These tests are assumed to tap various aspects of executive functioning such as planning, problem solving, reasoning, and response inhibition. Frequently used test batteries are the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001; Delis, Kramer, Kaplan, & Holdnack, 2004; Homack, Lee, & Riccio, 2005) and the Behavioural Assessment of the Dysexecutive Syndrome (BADS: Wilson et al., 1996). The BADS covers several facets of the dysexecutive syndrome, such as planning, problem-solving, monitoring, set shifting, etc. The BADS claims to be “a test battery aimed at predicting everyday problems arising from the dysexecutive syndrome” (Wilson et al., 1996). However, this superiority of the BADS compared with other executive tests to predict executive competency in daily life seems limited (Manchester, Priestley & Jackson, 2004; Norris & Tate, 2000).

Many of the tests are designed and provide normative data for severe executive deficits, or mainly for frontal dysfunction (e.g., the Wisconsin Card Sorting Test, Milner, 1963; Stroop Test, Perret, 1974; Tower of London, Shallice, 1982). However, their capacity to quantify milder executive impairments or the consequences of brain damage that is not restricted to the frontal areas is not well established. The patients of the present study experienced substantial executive impairments in their personal, social and vocational lives. Nevertheless, their ability to benefit from treatment was not entirely threatened by the severity of their dysexecutive impairments. Also, their lesions were not confined to frontal areas of the brain and were of various aetiologies. The first aim of this study was to examine whether several executive questionnaires and tests were valid tools to assess the executive deficits, or mainly for frontal dysfunction (e.g., the Wisconsin Card Sorting Test, Milner, 1963; Stroop Test, Perret, 1974; Tower of London, Shallice, 1982). However, their capacity to quantify milder executive impairments or the consequences of brain damage that is not restricted to the frontal areas is not well established. The patients of the present study experienced substantial executive impairments in their personal, social and vocational lives. Nevertheless, their ability to benefit from treatment was not entirely threatened by the severity of their dysexecutive impairments. Also, their lesions were not confined to frontal areas of the brain and were of various aetiologies. The first aim of this study was to examine whether several executive questionnaires and tests were valid tools to assess the executive complaints of this mixed group of brain-injured patients participating in an outpatient rehabilitation program.

A second aim of the study was to examine which of the used executive measures adequately discriminate between this specific group of brain-injured subjects and non-brain-injured subjects. The assessment of executive dysfunction often requires the use of an extensive array of instruments. This is obviously a time-consuming pursuit, and viewed in the light of the costs and the burden for the patient, it is desirable to find out if the number of assessment tools can be reduced. Also, as the number of tests administered increases, so does the likelihood of false-positive results (Burgess, 2003). We investigated which set of executive measures could effectively differentiate the performance of the brain-injured subjects from a group of non-brain-injured subjects, in order to find a more parsimonious way of assessing executive problems in this outpatient population.

Method

Subjects

Eighty-one patients participated in this study. They were all recruited from seven Dutch outpatient rehabilitation clinics and all of them had documented neurological injury. There were 34 patients with traumatic brain injury (TBI), 18 with a left hemisphere stroke, 13 with right hemisphere stroke, 4 with subcortical strokes, and 12 with different aetiologies (tumour, encephalitis, hypoxia). Considering the mixed nature of the aetiology of the brain injury in the subjects, executive deficits could vary in nature and severity. All patients were in the chronic stage of their disease, in which no spontaneous recovery was expected. The time since injury at which assessment took place was at least 6 months (median = 20 months). Duration of post-traumatic amnesia (PTA) in the TBI group was available for 21 persons (median = 16 days). Mean age was 42.83 years (SD = 13.75) and the female/male ratio was 26/55. Intelligence levels were established by a shortened version of a Dutch IQ test, the Groninger Intelligence Test (Luteijn & van der Ploeg, 1983), which showed a mean IQ of 112.55 (SD = 17.32), based on three non-paced subtasks.

A control group of 57 non-brain-damaged subjects participated in the study. There were 34 females and 23 males in this group. Mean age was 47.84 years (SD = 11.43). The mean IQ of this group was 120.93 (SD = 14.81). Educational levels were established by a commonly used Dutch classification system (Verhage, 1964). Educational levels were the same in both groups (median = 5.0) and although there was a difference between the groups with respect to age, a t-test revealed that this was not significant. Control subjects and patients were not perfectly matched with regard to gender distribution, however, there are no indications that gender may significantly influence executive functioning. Despite a significant difference in IQ scores between the groups (p < .05), no influence of this variable on neuropsychological test scores (see below) could be found with an analysis of covariance.

Procedure

The selection of patients was based on the subjective complaints of the patients and the observations by proxies of executive dysfunction in daily life. After referral for outpatient rehabilitation, patients and one of their proxies were interviewed by a neuropsychologist. In this interview a checklist was used that covered the main aspects of executive dysfunction (Spikman, 2002). This checklist contributed to the confirmation that the reported
executive problems were the most prominent complaints. Indicative for executive dysfunction were difficulties with planning and organizing everyday tasks and difficulties with problem-solving and reasoning, as well as symptoms of distractibility, impulsivity and disinhibition.

Exclusion criteria were the following: progressive neurological disease, severe behavioural disorders, severe verbal, sexual or aggressive disinhibition, severe loss of initiative (abulia), a total lack of awareness of deficit (anosognosia), severe aphasia, severe hemi-inattention, previous psychiatric disorders, and substance abuse.

Upon recruitment, the patients were given a set of typical ‘executive’ tests and a number of relevant questionnaires. To ascertain that the patients’ complaints concerned mainly executive difficulties rather than other consequences of brain damage two memory tests were administered: The 15 Words Test, a Dutch version of the Rey Auditory Verbal Learning Test (Saan & Deelman, 1986) and the Dutch translation of Digits Forward (DF) and Digits Backward (DB) from the Wechsler Adult Intelligence Scale (WAIS-III, Wechsler, 2000). The performances of our patient group in the Digit Span Test (mean DF = 7.6, SD = 2.5 and mean DB = 5.5, SD = 1.9) fell within the normal range taken from the WAIS III test manual (mean DF = 6.1, SD = 1.0 and mean DB = 4.9, SD = 1.4). Comparison of the performances of the patients with normative data for the 15 Words Test, provided by Hochstenbach, Mulder, van Limbeek, Donders, and Schoonderwaldt (1998), showed no substantial differences between the groups: Immediate recall controls/patients 43.79/38.93, SD = 12.04/10.89; Delayed recall controls/patients 28.47 /27.21, SD = 3.43/3.44; Recognition controls/patients 9.76/7.79, SD = 3.43/3.44; Recognition controls/patients 28.47/27.21, SD = 3.13/4.11.

Questionnaires
A selection of questionnaires and rating scales was used, with the purpose of assessing executive problems in daily life as close as possible. The first questionnaire used was the Dysexecutive Questionnaire (DEX; Burgess et al., 1996). One version of the questionnaire was completed by the patient (DEX-self), another was completed by a proxy, i.e., someone who knew the patient well (usually the partner) (DEX-PR). A third version was completed by a professional (usually a neuropsychologist), who had interviewed the patient and his/her proxy, but was not further involved in this study (DEX-TH). Thus, the degree of the patients’ executive impairments was also judged by two independent raters.

A second inventory, the Executive Observation Scale (EOS; Pollens, McBratnie, & Burton, 1988) consists of 8 items covering both cognitive and behavioural aspects of executive functioning in everyday tasks (awareness, planning, goal setting, self-initiation, self-inhibition, self-monitoring, ability to change set, strategic behaviour). Each of the items is rated on a Likert-type scale from 1 (complete inability) to 4 (completely independent and able). The EOS was completed by the same therapist mentioned above, who was not further involved in this study.

Besides the DEX and EOS, several other questionnaires and rating scales, partially tapping executive problems, were administered in order to get a comprehensive view on the patients’ daily life functioning. Both patients and proxies were asked to fill in the Patient Competency Rating Scale (PCRS; Prigatano et al., 1996). This is a 30-item self-report questionnaire that was originally designed to assess awareness of deficits in post-acute patients following brain injury. It asks patients and family members to make independent judgements about the perceived degree of competency the patient demonstrates on various behavioural, cognitive and emotional tasks. Judgements are based on the perception of how easy or difficult a certain behavioural activity presently is by choosing one of five responses: ‘can do with ease’, ‘fairly easy to do’, ‘with some difficulty’, ‘very difficult to do’, and ‘can’t do’.

The Apathy Evaluation Scale (AES; Marin, Biedrzycki & Firinciogullari, 1991) was completed by both patients and proxies. This questionnaire includes 18 assertions concerning the level of self-initiative and apathy. Responses are rated on a Likert-type scale, ranging from 1 (not true) to 4 (true).

Only the proxies were asked to mark one of three possible answers on the Behavioural and Emotional Change Questionnaire (BECQ; Spikman & Brouwer, 2002). This Dutch questionnaire is designed to gather information on behavioural and emotional changes after brain injury and rates how often certain behaviour is observed compared to pre-injury (‘equally’, ‘more often/worse’ and ‘much more often/worse’).

In order to measure executive functioning at participation level, the Role Resumption List (RRL; Spikman, Brand, & Brouwer, 2002) was administered. This is a structured interview, making an inventory of four areas of everyday life (vocational functioning, leisure activities, social interaction with family, and mobility). Each of these four areas is rated on a five-point scale by a therapist (usually a neuropsychologist), judging whether and how much the patients’ role-behaviour has changed since their brain injury. A score of 5 represents no change; a score of 1 represents a severe loss of independence compared to the time before the brain injury.

Finally, quality of life was measured with the Quality of Life after Brain Injury (QOLIBRI; Von Steinbuechel, Petersen, & Bullinger, 2005). This
would perform everyday activities. These activities differed on two dimensions: (1) the structure or the degree to which they were ‘open’ versus ‘closed’, and (2) the frequency of performing them. From this task the total number of relevant steps generated by the subjects was counted (EDTrel) (see Dritschel et al., 1998). We added two test variables to the original scoring system: the total number of irrelevant steps (EDTirr) and the proportion of irrelevant steps compared to the total number of steps (EDTratio).

Administration
The above-mentioned tests and questionnaires were administered to the patients in two or three sessions. The total time to complete both tests and questionnaires took approximately four to five hours. The patients’ proxies needed approximately half an hour to complete the questionnaires. All of the tests were also administered to the control subjects, as well as the DEX. Control subjects were not eligible for treatment, so their DEX was only completed by themselves (DEX-self) and their proxies (DEX-PR) and not by a therapist.

Statistical analyses
Due to skewness of the data, both in the patient sample as well as in the sample of healthy controls, nonparametric statistical tests were applied. Spearman rank-order correlations were calculated as measures of association. Mann-Whitney U tests were conducted to assess group differences. Logistic regression was used as no assumptions have to be met on the normality of the independent variables (Menard, 2005).

Results

Questionnaires
DEX total scores were calculated by adding up the scores of each of the 20 items to create an overall measure of experienced or observed severity of executive impairment. This was done for the patients’ ratings (N = 81, mean = 31.93, SD = 13.56), proxies’ ratings (N = 78, mean = 31.77, SD = 14.88) and for the therapists’ ratings (N = 81, mean = 34.88, SD = 12.30) separately. A higher score represents greater impairment.

The DEX self-ratings of patients were compared to those of a group of healthy control subjects (N = 53, mean = 18.28, SD = 8.61). A Mann-Whitney U test for independent samples revealed a significant difference between the

Neuropsychological tests
In addition to the behavioural measures, a test battery of executive functioning was administered to the patients. The tests were chosen to represent various aspects of executive functions, such as planning abilities, flexibility, reasoning, problem-solving abilities and inhibition. Most of the tests were chosen on pragmatic grounds because they are commonly used in The Netherlands. Two of the tests are less well known, the Twenty Questions Test (Laine & Butters, 1982) and the Everyday Descriptions Task (EDT; Dritschel, Kogan, Burton, Burton, & Goddard, 1998). These were chosen because they are more ‘open-ended’ and ‘ill-structured’ tasks. The BADS was chosen because it is known as a more ‘ecologically valid’ test of executive functioning. The following variables were used in this study:

- Trail Making Test (Reitan, 1958), ratio of time to complete part B compared to the time to complete part A (TMT B/A).
- Stroop Test (Stroop, 1935), ratio of time to complete part three compared to the time to complete part two (STR 3/2).
- Verbal Fluency: animals and professions (from the Groninger Intelligence Test; Luteijn & van der Ploeg, 1983), total fluency score (VF).
- Tower of London (Shallice, 1982), total number of correct solutions (TOL).
- Behavioural Assessment of the Dysexecutive Syndrome was used (Dutch translation; Krabbendam & Kalff, 1997): the Rule Shift Card Test number of errors (RS), the Action Program Test raw score (AP), the Key Search Test raw score (KS), the Temporal Judgement Test raw score (TJ), the Zoo Map Test total raw score (ZOO), and the Modified Six Elements Test total raw score (MSET).
- Twenty Questions Test (Laine and Butters, 1982): the subjects were asked to guess which of 42 drawings of objects representing overlapping classes, such as animals, clothing, or round objects, the examiner had in mind, by asking questions, to which the examiner only answered with ‘yes’ or ‘no’. A maximum of 20 questions was allowed before the test was ended. The number of questions asked is taken as the variable of interest here (20Q).
- A translation of the Everyday Descriptions Task (Dritschel et al., 1998) consisted of eight questions requiring the subjects to describe how they
self-ratings of patients and healthy controls ($z = -5.743, p = .000$). A significant difference was also found between DEX proxy ratings of patients (N = 78, mean = 31.27, SD = 14.88) and controls (N = 49, mean = 18.10, SD = 9.93, $z = -5.11, p = .000$). This suggests that this specific group of patients had significantly more executive problems in everyday life than healthy controls, not only as self-observed, but also as judged by proxies.

An analysis of variance using a repeated measures design was carried out with the three DEX ratings, in which the assumption of sphericity was met. The results showed that the within-subjects factor ‘DEX rater’ was not significant, $F(2, 154) = 2.439, p = .091$. In other words, no significant differences in mean DEX total scores were found between the three observers. It was not necessary to conduct a non-parametric analysis, because a repeated measures design is not sensitive to skewness (Rietveld & van Hout, 2007) and provides greater power to detect effects than a Friedman test.

Correlations computed between the DEX ratings of patients and proxies ($r = .489, p < .01$), patients and therapists ($r = .425, p < .01$), and proxies and therapists ($r = .405, p < .01$) were significant, but modest, explaining between 16% and 24% of the variance.

To investigate the concurrent validity of behavioural measures of executive functioning, correlations between DEX total scores of the patients, proxies and therapists and scores from the nine other questionnaires were established. The correlation matrix is presented in Table 1.

Considering the correlations in Table 1, significant correlations were found between total DEX scores of patients and the majority of the questionnaires, except for the EOS and the RRL. Significant correlations were also found between total DEX scores of proxies and all of the questionnaires, except for the EOS. Correlations between total DEX scores of therapists and the other questionnaires were weaker. Five out of nine questionnaires correlate significantly with the DEX ratings of therapists. Thus, the presence of executive impairments was not only apparent in the total DEX scores, but was also reflected in several other questionnaires that measure different, but obviously related, aspects of daily life functioning.

Inspection of correlations among the nine other rating scales (not illustrated here) reveals an overall correspondence in reported problems across rating scales and raters. Mutual correlations are all significant, except for the EOS, which only correlates significantly with the RRL and the BECQ.

In summary, significant, but rather low mutual correlations between all the questionnaires were found within the three rater groups. Between groups, the strongest associations were found between patients’ and proxies’ ratings, whereas weaker associations were found between therapists’ and patients’ scores and therapists’ and proxies’ scores. In particular, the EOS correlated poorly with the other questionnaires.

### Table 1 Spearman rank-order correlation coefficients between DEX ratings and Questionnaires

<table>
<thead>
<tr>
<th></th>
<th>DEX-self (N = 81)</th>
<th>DEX-PR (N = 78)</th>
<th>DEX-TH (N = 81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QOLIBRI (satisfaction)</td>
<td>-.539**</td>
<td>-.469**</td>
<td>-.256*</td>
</tr>
<tr>
<td>QOLIBRI 2 (burden)</td>
<td>.602**</td>
<td>.335**</td>
<td>.253*</td>
</tr>
<tr>
<td>AES patient (N=81)</td>
<td>-.368**</td>
<td>-.312**</td>
<td>-.084</td>
</tr>
<tr>
<td>AES proxy (N=78)</td>
<td>-.295**</td>
<td>-.518**</td>
<td>-.200</td>
</tr>
<tr>
<td>PCRS patient (N=81)</td>
<td>-.609**</td>
<td>-.575**</td>
<td>-.405**</td>
</tr>
<tr>
<td>PCRS proxy (N=78)</td>
<td>-.254*</td>
<td>-.713**</td>
<td>-.154</td>
</tr>
<tr>
<td>BECQ proxy (N=76)</td>
<td>.305**</td>
<td>.583**</td>
<td>.167</td>
</tr>
<tr>
<td>EOS (N=80)</td>
<td>-.088</td>
<td>-.141</td>
<td>-.480**</td>
</tr>
<tr>
<td>RRL (N=74)</td>
<td>.034</td>
<td>.478**</td>
<td>.252*</td>
</tr>
</tbody>
</table>

Note. QOLIBRI = Quality of Life after Brain Injury; AES = Apathy Evaluation Scale; PCRS = Patient Competency Rating Scale; BECQ = Behavioural and Emotional Change Questionnaire; EOS = Executive Observation Scale; RRL = Role Resumption List.

* Correlation significant at the 0.05 level (2-tailed).
** Correlation significant at the 0.01 level (2-tailed)
To determine whether performance of patients on the BADS and other neuropsychological tests were associated with their scores on the DEX, the three DEX ratings (patient, proxy and therapist) were correlated with all test variables (Table 3). Correlations between the DEX ratings and test scores were rather low and did not attain statistical significance in most cases. Significant correlations were only found between DEX-self and BADS RS ($r = .274$, $p < .05$), DEX-TH and BADS RS ($r = .245$, $p < .05$), DEX-TH and EDTratio ($r = .245$, $p < .05$), between TOL and both DEX-self ($r = -.236$, $p < .05$) and DEX-PR ($r = -.253$, $p < .05$).

### Table 2
Mean raw scores and (Standard Deviations), results of Mann-Whitney U tests (one-tailed), and effect sizes of group differences for patients (N=81) and healthy controls (N=57)

<table>
<thead>
<tr>
<th>Test</th>
<th>Patient Mean (SD)</th>
<th>Control Mean (SD)</th>
<th>z and p (1-tailed)</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BADS RS</td>
<td>1.89 (2.96)</td>
<td>1.12 (2.26)</td>
<td>z = -.2.891, p = .002</td>
<td>.28</td>
</tr>
<tr>
<td>BADS AP</td>
<td>4.60 (.85)</td>
<td>4.75 (.87)</td>
<td>z = -.2.031, p = .021</td>
<td>.17</td>
</tr>
<tr>
<td>BADS KS</td>
<td>11.40 (3.65)</td>
<td>12.93 (2.33)</td>
<td>z = -.2.612, p = .0045</td>
<td>.44</td>
</tr>
<tr>
<td>BADS TJ</td>
<td>2.11 (.85)</td>
<td>2.33 (.95)</td>
<td>z = -.1.381, p = .114</td>
<td>.25</td>
</tr>
<tr>
<td>BADS MSET</td>
<td>4.99 (.136)</td>
<td>5.71 (.72)</td>
<td>z = -.3.202, p = .0005</td>
<td>.63</td>
</tr>
<tr>
<td>BADS ZOO</td>
<td>8.28 (4.27)</td>
<td>11.54 (3.79)</td>
<td>z = -.4.031, p = .000</td>
<td>.80</td>
</tr>
<tr>
<td>EDTrel</td>
<td>37.59 (13.98)</td>
<td>43.19 (16.93)</td>
<td>z = -.1.908, p = .028</td>
<td>.34</td>
</tr>
<tr>
<td>EDTirr</td>
<td>6.73 (9.41)</td>
<td>2.85 (2.71)</td>
<td>z = -.3.416, p = .0005</td>
<td>.51</td>
</tr>
<tr>
<td>EDTratio</td>
<td>.14 (12)</td>
<td>.06 (.06)</td>
<td>z = -.4.193, p = .000</td>
<td>.76</td>
</tr>
<tr>
<td>20Q</td>
<td>13.44 (5.22)</td>
<td>8.96 (4.42)</td>
<td>z = -.4.974, p = .000</td>
<td>.91</td>
</tr>
<tr>
<td>STR 3/2</td>
<td>1.61 (.36)</td>
<td>1.49 (.20)</td>
<td>z = -.2.651, p = .004</td>
<td>.38</td>
</tr>
<tr>
<td>TMT B/A</td>
<td>2.25 (.86)</td>
<td>2.37 (.83)</td>
<td>z = -.1.119, p = .131</td>
<td>.15</td>
</tr>
<tr>
<td>VF</td>
<td>36.63 (10.30)</td>
<td>43.13 (10.88)</td>
<td>z = -.3.575, p = .000</td>
<td>.62</td>
</tr>
<tr>
<td>TOL</td>
<td>10.69 (1.44)</td>
<td>11.41 (.91)</td>
<td>z = -.3.616, p = .000</td>
<td>.57</td>
</tr>
</tbody>
</table>

**Note.** BADS = Behavioural Assessment of the Dysexecutive Syndrome; EDT = Everyday Descriptions Task; 20Q = 20 Questions Test; STR = Stroop test; TMT = Trail Making Test; VF = Verbal Fluency; TOL = Tower of London test.

### Table 3
Spearman rank-order correlation coefficients between patients’ raw test scores (N=81) and DEX total scores

<table>
<thead>
<tr>
<th></th>
<th>DEX-self (N = 81)</th>
<th>DEX-PR (N = 78)</th>
<th>DEX-TH (N = 81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BADS RS</td>
<td>.274*</td>
<td>.245*</td>
<td>-.044</td>
</tr>
<tr>
<td>BADS AP</td>
<td>-.087</td>
<td>-.133</td>
<td>-.137</td>
</tr>
<tr>
<td>BADS KS</td>
<td>-.019</td>
<td>-.048</td>
<td>-.042</td>
</tr>
<tr>
<td>BADS TJ</td>
<td>.097</td>
<td>-.036</td>
<td>-.042</td>
</tr>
<tr>
<td>BADS MSET</td>
<td>-.015</td>
<td>-.020</td>
<td>-.161</td>
</tr>
<tr>
<td>BADS ZOO</td>
<td>.192</td>
<td>.072</td>
<td>-.014</td>
</tr>
<tr>
<td>EDTrel</td>
<td>-.129</td>
<td>-.156</td>
<td>-.170</td>
</tr>
<tr>
<td>EDTirr</td>
<td>.012</td>
<td>-.055</td>
<td>.196</td>
</tr>
<tr>
<td>EDTratio</td>
<td>.049</td>
<td>-.009</td>
<td>.245*</td>
</tr>
<tr>
<td>20Q</td>
<td>.053</td>
<td>.055</td>
<td>.105</td>
</tr>
<tr>
<td>STR 3/2</td>
<td>.094</td>
<td>.049</td>
<td>.076</td>
</tr>
<tr>
<td>TMT B/A</td>
<td>.009</td>
<td>-.166</td>
<td>-.090</td>
</tr>
<tr>
<td>VF</td>
<td>.187</td>
<td>-.151</td>
<td>.132</td>
</tr>
<tr>
<td>TOL</td>
<td>-.236*</td>
<td>-.253*</td>
<td>-.113</td>
</tr>
</tbody>
</table>

**Note.** DEX = Dysexecutive Questionnaire; BADS = Behavioural Assessment of the Dysexecutive Syndrome; EDT = Everyday Descriptions Task; 20Q = 20 Questions Test; STR = Stroop test; TMT = Trail Making Test; VF = Verbal Fluency; TOL = Tower of London test.

* Correlation significant at the 0.05 level (2-tailed).

** Correlation significant at the 0.01 level (2-tailed).
Prediction of group membership

In order to examine to what extent the DEX ratings, the executive test measures or a combination of tests and DEX ratings were predictive of group membership (brain-injured or non-brain-injured), three sequential stepwise logistic regression analyses were conducted. In the first logistic regression analysis with DEX ratings only, the DEX ratings of patients and their proxies and the DEX self-ratings of healthy controls and their proxies were entered together as potential predictors of group membership (Table 4). The analysis revealed that the combined variables were a significant predictor of group membership ($\chi^2 = 40.293$, $df = 2$, $p < .000$), correctly classifying 79.5% of the patients and 65.3% of the control subjects. The overall classification rate was 74%.

Table 4 Results of logistic regression using DEX ratings as predictor variables of group membership (patients vs. controls)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>Wald</th>
<th>Significance</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEX-self</td>
<td>-.066</td>
<td>8.443</td>
<td>.004</td>
<td>.936</td>
</tr>
<tr>
<td>DEX-PR</td>
<td>-.049</td>
<td>5.231</td>
<td>.022</td>
<td>.952</td>
</tr>
</tbody>
</table>

Note. DEX = Dysexecutive Questionnaire

The results of the second logistic regression analysis using all the executive test variables as potential predictors, provided a statistically reliable model ($\chi^2 = 63.415$, $df = 6$, $p < .000$) in which VF, EDTratio, 20Q, BADS KS, BADS MSET, and BADS ZOO contributed significantly to the classification of the groups (Table 5). Correct classification of patients was good (87.5%), but classification success of control subjects was poorer (69.8%). Overall classification success was good: 81.3%.

Table 5 Results of logistic regression using test variables as predictor variables of group membership (patients vs. controls)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>Wald</th>
<th>Significance</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>VF</td>
<td>.066</td>
<td>6.433</td>
<td>.011</td>
<td>1.068</td>
</tr>
<tr>
<td>EDTratio</td>
<td>-.132</td>
<td>9.381</td>
<td>.002</td>
<td>.000</td>
</tr>
<tr>
<td>20Q</td>
<td>-.135</td>
<td>6.254</td>
<td>.012</td>
<td>.874</td>
</tr>
<tr>
<td>BADS KS</td>
<td>.161</td>
<td>4.741</td>
<td>.029</td>
<td>1.175</td>
</tr>
<tr>
<td>BADS MSET</td>
<td>.880</td>
<td>6.542</td>
<td>.011</td>
<td>2.411</td>
</tr>
<tr>
<td>BADS ZOO</td>
<td>.141</td>
<td>4.250</td>
<td>.039</td>
<td>1.152</td>
</tr>
</tbody>
</table>

Note. VF = Verbal Fluency; EDT = Everyday Descriptions Task; 20Q = 20 Questions Test; BADS = Behavioural Assessment of the Dysexecutive Syndrome.

Results of the third logistic regression analysis, which contained both tests and DEX ratings (self and proxy) as potential predictors, demonstrated that the DEX-self, the VF, the EDTratio, and the 20Q contributed significantly to the classification of groups ($\chi^2 = 78.068$, $df = 6$, $p < .000$; Table 6). As in the previous analyses, correct classification of patients was higher (90.9%) than of controls (79.5%). The overall percentage of predicted classification was high at 87.1%.

Table 6 Results of logistic regression using test variables and DEX ratings as predictor variables of group membership (patients vs. controls)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>Wald</th>
<th>Significance</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>VF</td>
<td>.092</td>
<td>7.338</td>
<td>.007</td>
<td>1.096</td>
</tr>
<tr>
<td>EDTratio</td>
<td>-.135</td>
<td>6.488</td>
<td>.011</td>
<td>.000</td>
</tr>
<tr>
<td>20Q</td>
<td>-.232</td>
<td>10.410</td>
<td>.001</td>
<td>.793</td>
</tr>
<tr>
<td>BADS KS</td>
<td>.181</td>
<td>3.787</td>
<td>n.s.</td>
<td>1.198</td>
</tr>
<tr>
<td>BADS MSET</td>
<td>.742</td>
<td>3.646</td>
<td>n.s.</td>
<td>2.099</td>
</tr>
<tr>
<td>DEX-self</td>
<td>-.115</td>
<td>12.619</td>
<td>.000</td>
<td>.892</td>
</tr>
</tbody>
</table>

Note. VF = Verbal Fluency; EDT = Everyday Descriptions Task; 20Q = 20 Questions Test; BADS = Behavioural Assessment of the Dysexecutive Syndrome; DEX = Dysexecutive Questionnaire.
Discussion

In this study the executive impairments of a mixed group of brain-injured patients referred for outpatient rehabilitation, were investigated. The patients were selected on the basis of subjective complaints and observations by proxies of executive dysfunction in daily life. Subsequently, they were assessed with a series of questionnaires and neuropsychological tests. The following main question was addressed: How are the executive problems experienced in everyday life situations in this particular group of patients reflected in behavioural measures and neuropsychological tests?

The results showed that self-ratings of patients on the DEX Questionnaire were significantly higher than those of healthy control subjects, which confirms that this specific group of patients indeed experiences considerable executive problems in everyday life. There were no significant differences found between mean total DEX scores of patients, proxies and therapists, which confirms the presence of substantial executive problems in the patient group under study, but is in contrast with earlier findings (Bennett et al., 2005a; Burgess et al., 1998; Knight et al., 2002; Wilson et al., 1996), in which patients’ scores on the DEX Questionnaire turned out to be substantially lower than those of independent raters (i.e., carers or therapists). The discrepancy between problems experienced by patients and those observed by others found in these studies was interpreted as an indication of lack of insight. Lack of insight or awareness is often observed in persons with executive deficits and appears to be related to damage to the prefrontal regions of the brain (Stuss, 1991). Apparently, patients in our study could estimate their executive problems as accurately as their independent observers. This may indicate that the current group of patients showed relative intact awareness into their executive problems. This level of awareness might be explained by the patients’ long sojourn at home, in which they were repeatedly confronted with daily executive problems. This problem experience might have increased their awareness. The relatively intact awareness of executive deficits might also be explained by the residual learning capacities possessed by the brain-injured patients of this study. Their eligibility for outpatient rehabilitation indicates reasonable residual learning capacities and, apparently, better awareness than the patients participating in the above-mentioned studies.

While correlations between DEX ratings of patients, proxies and therapists were statistically significant, they were nevertheless modest. This suggests that the three raters may have been inaccurate or inconsistent in observing everyday executive problems. As Norris and Tate (2000), for example,
observed about low inter-rater reliability in DEX ratings “the close relatives of the participants completing the DEX proved to be extremely heterogeneous. It was difficult to be confident that all relatives were completing the DEX with similar degrees of awareness, understanding and appreciation of the participant’s everyday functioning”.

In order to establish the concurrent validity of the DEX in comparison with several other behavioural indicators of executive impairment, the DEX scores were correlated with the scores of nine other questionnaires, some of which were also filled in by proxies and therapists. Significant mutual correlations suggest an underlying common factor that represents executive dysfunction when assessed with different rating scales. These results suggest that the DEX possesses adequate concurrent validity and can be considered as a valid questionnaire of executive functioning in this particular group of brain-injured patients.

Comparison of the performances of patients and control subjects on 14 neuropsychological test measures revealed that patients performed significantly worse on the majority of these measures. Thus, executive tests appeared to be sufficiently sensitive in identifying executive deficits in these patients. However, the correlations between the DEX ratings and test scores were generally low. The tests did not predict presence and severity of everyday executive problems as measured by the DEX.

An explanation for this lack of association between DEX-ratings and test performances might be that the tests are intended to measure cognitive aspects of executive functioning, such as ‘planning’ or ‘flexibility’, whereas questionnaires such as the DEX are designed to cover a much broader range of symptoms including emotional, behavioural and motivational aspects of executive dysfunction. Moreover, because tests are intended to measure executive functioning at a cognitive level, while questionnaires target executive functioning at a cognitive-behavioural level in everyday life tasks, strong correlations between the two cannot be expected. As previously mentioned, DEX ratings might have been seriously inaccurate, which also explains the finding of low correlations between DEX ratings and test scores.

Specifically, correlations between DEX self-observations of patients and test performances were not significant, except for those between the DEX-self and the BADS RS, as well as the TOL, which were significant, but low. Previous studies have reported similar results. Wilson et al. (1996) found that none of the BADS subtests or total profile score correlated with the patients’ own ratings of their dysexecutive problems. Burgess et al. (1998) also found that none of the executive test measures used in their study were significantly correlated with DEX ratings of patients. Moreover, Bennett et al. (2005a) reported that BADS subtests were not significantly correlated with patients’ DEX ratings. Finally, Evans, Chua, McKenna, and Wilson (1997) found no correlations between any of the BADS subtest scores and self-ratings on the DEX for a schizophrenic or a brain-injured group.

Two significant correlations between proxies’ DEX ratings and test scores of patients were found (BADS RS and TOL). The results of earlier studies were not unequivocal. Few or no significant correlations between relatives’ DEX ratings and BADS subtests and/or other tests of executive functioning were found in Bennett et al., 2005a; Evans et al., 1997; Norris & Tate, 2000, while others reported significant correlations between test performances and DEX-ratings (Burgess et al., 1998; Wilson et al., 1996).

Knight et al. (2002) found significant correlations between DEX ratings of staff members and the BADS Zoo Map Test and error scores of the Multiple Errands Test. Bennett and colleagues (2005a, Bennet, Ong, & Ponsford, 2005b) found several significant correlations between the therapist’ DEX and tests. In our study only the EDT correlated significantly with DEX ratings of therapists. An explanation for these divergent results might be that the patients in the Bennett study were observed for longer periods of time before the DEX was completed (ranging from two weeks to several months) and were seen at least three times by each professional, whereas in our study the DEX was in some instances filled in after a short acquaintance with the patients. The DEX-ratings of therapists in the Bennett study may therefore have been more dependable and have allowed the authors to find the mentioned results. Also, the EOS’s weak correlation with the DEX patient and DEX proxy is another indication that the therapists in our study may have been inaccurate in their ratings.

A second aim of this study was to investigate which measure or combination of measures could optimally discriminate between this specific outpatient group and non-brain-injured subjects. This objective of reducing the number of executive assessment tools is worth trying, in order to reduce assessment costs and burden for the patients. Moreover, it is well known that tests of executive function do not always discriminate adequately between clinical and non-clinical samples (Andres & van der Linden, 2001; Burgess et al., 1998; Manchester, 2004). Especially in ‘higher functioning’ patients with milder or more subtle degrees of executive impairment (Norris & Tate, 2000; Rath, Simon, Langenbahn, Sherr, & Diller, 2000; Rath, Langenbahn, Simon, Sherr, & Diller, 2004; von Cramon & Matthes-von Cramon, 1994) it can be expected that the tests’ ability to reliably discriminate patients from non-brain-injured controls is attenuated. Therefore, we examined to which
A first logistic regression analysis showed that the DEX-self and DEX-PR proxy as a set contributed to an adequate classification in the patient group (79.5%) and the control group (65.3%). The second logistic regression analysis revealed that by using test variables as predictors the VF, the EDT, the 20Q, the BADS KS, BADS MSET, and the BADS ZOO could correctly discriminate the patient group (87.5%) from the control group (69.8%). Finally, the combination of both the DEX ratings and test variables was most accurate in discriminating patients from healthy control participants. The DEX-self, the VF, the 20Q, and the EDT made significant contributions to the prediction of group membership, correctly classifying 90.9% of the patients and nearly 80% of the controls (79.5%). Norris and Tate (2000), who used the subtests from the BADS and other tests of executive function, but no DEX ratings, in order to discriminate neurological patients from healthy controls, correctly classified 64% of the patients and 83% of the controls. In comparison, our results show that the addition of DEX ratings increases the ability to detect patients. The use of a questionnaire, such as the DEX, maximizes the discriminative capacity of the instruments by nearly 30%, at a negligible expense of accuracy in the control group (approximately 3%). It appears that to discriminate this group of patients from healthy controls, the assessment of subjective complaints combined with ‘open-ended’ tasks provides the most predictive set of instruments. These open-ended tasks represent free-choice, ambiguous problem-solving situations of daily life better than the more structured traditional neuropsychological tests (Goldberg & Podell, 1999, 2000). The 10% difference in the predicted accuracy between patients and controls may be explained by the fact that dysexecutive-like behaviours may occur in non-brain-damaged individuals with a relative high frequency (Chan, 2001).

Although there was evidence that our patients did not have major memory deficits, it cannot be warranted that the only difference between patients and controls was in terms of executive deficits only. Hence, the used tests may have been sensitive to brain injury rather than executive functioning. Therefore, in further studies it might be interesting to include two clinical groups, one with evidence of executive impairment and one without, in order to predict membership in terms of executive impairments exclusively.

The problems discussed above point to the more general question of how neuropsychological questionnaires and tests should be validated. The main difficulty remains the absence of a real “gold standard” against which new assessment tools can be validated. What is more, many tests fail to predict real-world functioning and have little ecological validity (Burgess et al, 2006; Cripe, 1998). To improve the assessment of daily executive problems careful interviews with patients, their proxies and therapists, supplemented with valid questionnaires, are required. These interviews, combined with quantitative and structured observations of brain-injured patients with and without executive problems (and healthy controls) could be correlated with imaging data of the distributed neural network, comprising the frontal cortex, underlying executive functioning. This combination of anamnestic and hetero-anamnestic data, observations and neuroimaging data should be the keystone against which neuropsychological tests should be validated. It should also be noted that this kind of assessment of executive problems could benefit from a better input from cognitive science in specifying more accurately the wide variety of executive processes, in order to improve the construct validity of questionnaires and observation instruments as well as neuropsychological tests.

In summary, the presence and severity of executive deficits was established in a population with brain injury who were referred for outpatient rehabilitation because of executive problems, using both observational measures and neuropsychological tests. Correlations between observations and test performances were equally difficult to establish as in other cognitive domains (e.g. Azouvi et al., 2002; McGlone, 1994; Ponsford & Kinsella, 1991). However, a combination of observations and tests allows the clinician a fairly accurate identification of patients with executive problems. The results imply that a comprehensive description of the patients’ executive problems can be obtained in relatively short time and therefore in a more cost-effective way.
References


Spikman, J.M. (2002). Checklist Executive Stormissen [Checklist Executive Disorders]. Internal Publication University Medical Centre Groningen, the Netherlands.

Spikman, J.M., Brand, R., & Brouwer, W.H. (2002). Rolletrekkingslijst [Role Resumption List]. Internal Publication University Medical Centre Groningen, the Netherlands.

Stroom, J.R. (1903). Studies of interference in serial


Script generation in brain-injured patients with a dysexecutive syndrome

Submitted for publication
D.H.E. Boelen, P. Allain, J.M. Spikman, L. Fasotti
Abstract

This study investigates whether brain-injured patients with a dysexecutive syndrome have difficulties in script information processing. Forty-eight brain-injured patients were compared to 99 healthy control subjects, in a script generation task. Participants were asked to describe how to perform eight everyday activities. The script-items were manipulated so that they varied in structure (open-ended vs. closed) and in frequency of performing (high vs. low). Patients and control subjects evoked an equal number of actions, but patients produced significantly more irrelevant actions and made significantly more perseverative errors. They had the most problems in open-ended scripts and in low-frequency scripts. This investigation shows that not only patients with strictly prefrontal damage are impaired in script generation. The consequences for treatment are discussed.

Introduction

Planning a sequence of actions is regarded as one of the main functions of the frontal lobes (Shallice, 1988). According to Grafman (1989, 1994) the planning of complex, goal-directed behaviour depends on the activation of scripts. The term script refers to knowledge structures representing familiar activities, such as organizing a party or travelling by public transport. Grafman (1989, 1994) defines scripts as managerial knowledge units (MKUs), consisting of overlearned sequences of events, real or imaginary, that are retrieved automatically and that have a clear beginning and end. According to Grafman scripts are stored and activated in the frontal lobes, which are very important in the maintenance of the sequential order of events and fixed boundaries. Within Grafman's framework, damage to the prefrontal areas has a negative effect on the ability to retrieve and process script and on the cognitive representations of action knowledge as the MKUs themselves should be damaged.

More recently, Grafman (2002) and Wood and Grafman (2003), proposed an updated model of the original MKUs theory labelled Structured Event Complexes (SECs) framework. In this extension of the original model, SECs are goal-oriented sets of events that represent thematic knowledge, morals, abstractions, concepts, social rules, event features or grammars, and are structured in sequence. According to Wood and Grafman (2003), these representations are memories that are localized in neural networks that encode information, and, when activated, enable access to this stored information. This extension agrees with the MKUs model in proposing that contents and sequential characteristics of complex sequence of actions (scripts) are represented within the frontal lobe. Again, within the SECs model, frontal lobe damage might impair script processing, as scripts themselves are damaged.

Script processing has been investigated in studies either by means of script generation tasks (Cazalis, Azouvi, Sirigu, Agar, & Burnod, 2001; Chevignard et al., 2000; Dritschel, Kogan, Burton, Burton, & Goddard, 1998; Fortin, Godbout, & Braun, 2003; Godbout, Cloutier, Bouchard, Braun, & Gagnon, 2004; Godbout & Doyon, 1995; Godbout, Grenier, Braun, & Gagnon, 2005; Sirigu et al., 1995; Zanini, 2008; Zanini, Rumiati, & Shallice, 2002) or by script sorting and/or sequencing tasks (Allain et al., 1996; Allain, Etcharry-Bouyx, & Le Gall, 2001; Allain, et al., 2001; Allain, Le Gall, Etcharry-Bouyx, Aubin, & Emile, 1999; Sirigu et al., 1996; Zanini, 2008).

Specifically, script generation studies have shown that sequencing errors occur more often in patients with frontal lesions than in healthy control
subjects (Fortin et al., 2003; Godbout & Doyon, 1995; Godbout et al., 2004; Sirigu et al., 1995). Also, frontal lobe patients made more errors, such as irrelevant intrusions (Fortin et al., 2003) and boundary errors (Sirigu et al., 1995). Typically, all studies (except for Godbout et al., 2004) showed that frontal patients and healthy control subjects evoked approximately equal numbers of actions in script generation tasks (Fortin et al., 2003; Godbout et al., 2005; Sirigu et al., 1995; Zanini, Rumiai, & Shallice, 2002). The studies mentioned above have confirmed the sensitivity of script generation to frontal lobe damage, which disrupts the ability to activate coherent sequences of actions needed to achieve a goal, even for relatively familiar activities. Consistent findings come from the script sorting and sequencing studies mentioned before and from other studies examining patients with functional causes of reduced frontal activity such as Parkinson's disease (Godbout and Doyon, 2000; Zalla et al., 1998; Zalla et al., 2000). Huntington's disease (Allain et al., 2004), and normal ageing (Allain et al., 2007; Godbout, Doucet, & Fiola, 2000). Finally, neuroimaging studies with healthy subjects (Crozier et al., 1999; Knutson, Wood, & Grafton, 2004; Partiot, Grafton, Sadato, Flitman, & Wild, 1996; Ruby, Sirigu, & Decety, 2002) have shown that the frontal lobe is recruited for script processing.

However, with regard to script generation, two studies by Godbout and Doyon (1995) and Godbout et al. (2004) revealed that not only prefrontal regions are involved in script generation. The patients with frontal lesions in the study of 2004 evoked significantly fewer actions overall than patients with parietal lobe damage and healthy subjects. In both studies (1995, 2004), frontally damaged patients as well as patients with parietal lesions had significantly more difficulties with script sequencing than healthy control subjects. The results show that, although script generation has been primarily associated with prefrontal regions, (with a particular emphasis on their contribution in the sequencing of actions), other brain regions might be activated in script information processing. In particular, parietal information processing, or a “space-processing component” seems to be involved (Godbout et al., 2004, p868). Also, a study by Cazalis et al. (2001) showed that severe TBI patients without any focal cortical contusion on CT and/or MRI, produced an equal number of actions as healthy controls and had normal sequencing performance. However, these patients were significantly slower in evocation and made more closure errors, which was interpreted in terms of impaired information processing.

Two other studies cast doubts upon the idea that script processing is a specifically frontal function. First, an (fMRI) study conducted by Crozier et al. (1999) using scripts as stimuli, showed that action sequence processing is associated with activation of regions in the left frontal, parietal and temporal cortices. Second, Allain et al. (2008) have shown that patients with Alzheimer's disease, which are known to have progressive damage, predominantly affecting temporoparietal structures, were impaired in script sequencing and sorting tasks.

In the current study, we investigated whether patients with dysexecutive symptoms defined in a more functional sense (Baddeley & Wilson, 1988), had difficulties with script generation. Brain-damaged subjects with functional executive deficits constitute the majority of patients referred for rehabilitation, and script tasks might confirm that generating scripts involves a larger neural network than merely prefrontal tissue and can provide clues for the treatment of executive deficits in this category of patients. Therefore, in this study, we included brain-injured patients who complained of executive problems in daily life situations, or were observed by their proxies to have such problems (see also Boelen, Spikman, Rietveld, & Fasotti, 2009). These patients had brain injuries of various aetiologies, such as stroke, TBI, brain tumour, encephalitis and hypoxia. Thus, their injuries were often more diffuse and not only located in prefrontal regions. However, the included patients showed executive problems in neuropsychological tests and on the DEX questionnaire (Boelen et al., 2009). More specifically, we examined whether patients with a dysexecutive syndrome produce as many actions in a script generation task as healthy control subjects, just as shown in earlier studies (Cazalis et al., 2001; Dritschel et al., 1998; Fortin et al., 2003; Godbout et al., 2005; Sirigu et al., 1995; Zanini et al., 2002). Also, we explored if these patients have difficulties sequencing actions within generated scripts, just as patients with prefrontal lesions (Fortin et al., 2003; Godbout & Doyon, 1995; Godbout et al., 2004; Sirigu et al., 1995), or if they perform similarly to healthy control subjects (Cazalis et al., 2001; Zanini et al., 2002). Moreover, we wanted to know if these patients have problems with discarding irrelevant actions and made more perseverations than healthy control subjects. Furthermore, we investigated whether type of script was of influence on the number actions evoked by patients and healthy control subjects. Hence, scripts that differed in structure (open versus closed) and in frequency of execution (and thus in familiarity) were administered. We expected that patients with executive impairments would have more difficulties than control subjects with scripts that are more open in structure and/or low in frequency, because such scripts demand a higher degree of executive control from the subjects than scripts that are routinely executed (characterized by high frequency and/or closed in structure).
Method

Subjects
Forty-eight patients, recruited from two Dutch outpatient rehabilitation clinics, participated in this study. The patients were included on the basis of complaints (by themselves or observed by proxies) about everyday executive problems. Upon recruitment, patients and one of their proxies were interviewed by a neuropsychologist who used a checklist that covered the main aspects of executive dysfunction (Spikman, 2002). The aim of the checklist was to ascertain that executive problems were indeed the most prominent complaints after brain injury. Examples of executive problems were problems with reasoning, problem-solving and planning, difficulties with time-estimation, self-monitoring, and behavioural changes such as impulsivity or a lack of concern for social rules. Patients were excluded in case of progressive neurological disease, severe behavioural disorders, severe verbal, sexual or aggressive disinhibition, severe loss of initiative (abulia), a total lack of awareness of deficit (anosognosia), aphasia, severe hemi-inattention, previous psychiatric disorders, and substance abuse. The executive complaints were further assessed using tests usually considered as sensitive to the dysexecutive syndrome and the DEX-questionnaire (see table 2).

Thus, the patients were not included based on focal frontal lesions, like the patients in some previous studies on script generation. They had brain injury of various aetiologies (12 patients had a stroke in the right hemisphere, 11 patients had a stroke in the left hemisphere, 3 patients had subcortical strokes, 15 patients had TBI, and 7 patients had brain injury of resulting from other causes (tumour, encephalitis, or hypoxia)). Due to the mixed nature of the aetiology of brain injury in the participating subjects, executive deficits could vary in nature and severity. All patients were in the chronic stage of their disease, in which no spontaneous recovery was expected. In this patient sample the time since injury at which assessment took place was at least 6 months (median = 17.0 months; mean = 27.57, SD = 42.03). The duration of post-traumatic amnesia (PTA) of TBI patients was available for 13 subjects (mean = 21.39 days, median = 14.00 days). The mean age of the patients was 47.15 years (SD = 12.62). The male/female ratio was 32/16. Although 11 patients with a left-sided stroke were included, none of these suffered from aphasia. Patients with even slight aphasic disorders were excluded from the study.

A control group of 99 non-brain-damaged subjects was included in the study as a reference sample. Their mean age was 46.07 (SD = 12.46) and the male/female ratio was 45/54. Educational level was assessed with a seven point scale, following a commonly used Dutch classification system (Verhage, 1983). Educational levels of both groups were comparable (median = 5.0, range 3-7). Patients and control subjects were not matched with regard to gender distribution. However, there are no indications in the literature that gender may significantly influence executive functioning or script generation specifically.

Procedure
The script generation task was a free response task in which subjects had to structure their own responses. It was derived from the Everyday Description Task by Dritschel et al. (1998). Subjects were asked to describe how they would accomplish eight everyday activities: 1) purchase a household appliance, such as a washing machine, 2) take a trip with friends, 3) to get a new job, 4) organize a move, 5) do the washing up, 6) put on a shirt, 7) paint a room, 8) make a sand castle. The proposed scripts varied on two dimensions. First, they could be ‘closed’ (with only one way to accomplish the activity, where the actions involved in execution are well established, and usually executed over a short period of time) or ‘open’ (less structured activities with several ways of execution and loose connections between the actions comprising the activity, which can extend over a long period of time). The second dimension was frequency (low versus high frequency), representing the number of times a person is exposed to an event requiring the generation and execution of a script. Frequency was manipulated to ensure that the structure inherent in the script was not influenced by the frequency with which it is executed. Thus, there were four possible combinations: A) open - high frequency, B) open - low frequency, C) closed - high frequency, D) closed - low frequency. The open-low frequency scripts are less likely to have a ready-scripted plan for their execution.

Subjects were given instructions to verbally generate the sequence of actions corresponding to each activity in the correct order, from the beginning to the end. They were asked to recite all the actions needed and put them in the order in which they actually would occur. The eight activities were presented in the same order to all subjects. Patients were not prompted to continue to produce actions when they fell silent. Subjects’ responses were tape recorded, transcribed and scored. The norms used to score the scripts of patients were based on the scripts of the 99 normal control subjects.
Scoring
A normative study was conducted to establish a corpus of script actions for each script. Norms were based on the scripts of 99 healthy control subjects and were used to score the patients’ scripts. The scripts of the normal control subjects were rated by three independent raters. These raters registered a list of action frequencies for each of the eight activities. These lists were exchanged and judged by the other raters. Finally, a definite list of actions per script was put together which formed the corpus of the script. The scoring of the scripts was based on the methods of Allain (2000), Bower, Black, and Turner (1979), Dritschel et al. (1998), and Roman, Brownell, Potter, and Seibold (1987). The main issue in investigating the control subjects’ scripts was whether subjects agreed in the actions they mentioned to describe the script headings. We designated as the group’s stereotype or script those actions that were mentioned by more than a criterion percentage of subjects. Thus, to be included in the script corpus, an action had to be mentioned by at least 18% of the normal control subjects (see also Allain, 2000; Bower et al., 1979; Godbout & Doyon, 1995; Godbout et al., 2005; Roman et al., 1987). Actions that met these criteria were subsequently classified as major (mentioned by at least 60% of the normal control subjects), minor (mentioned by 40-59% of the normal control subjects), or trivial (mentioned by 18-39% of the normal control subjects). Thus, major actions are the set of elements judged to be the most central to the theme or goal of the target activity (e.g. "look for a job vacancy" and "write a letter" in the ‘new job’-script, or "compare prices" in the ‘household appliance’-script). Minor actions are those elements judged to be less critical than major elements, but nonetheless a proper part of the target activity (e.g. "rinse the dirt from the dishes before actually washing up"). Trivial actions are any element judged to be peripheral to (but still a part of) the target activity (e.g. unbuttoning a shirt before putting it on). Evoked actions that did not reach the minimal frequency of 18%, were marked as intrusions. Relevant intrusions belonged to the script (e.g. for the script ‘put on a shirt’ the action “tuck the shirt in to your trousers”). Irrelevant intrusions did not belong to the script (e.g. “choose a tie that matches with the shirt”, or a personalization, such as “if I wear blue trousers, I choose a shirt with a matching colour”) and were considered as errors. Thus, relevant actions consisted of major, minor, trivial actions and relevant intrusions. Errors, on the contrary, consisted of irrelevant intrusions, perseverations and sequencing errors.

An open task such as a script generation task leads to a large variability in the response styles of patients and control subjects, rendering the scoring procedure difficult. Several patients’ scripts and even normal control subjects’ productions were characterized by tangentiality and some showed a tendency to end their script productions prematurely. Some patients and control subjects produced highly personalized scripts. Sometimes, subjects used different formulations to denote the same actions (semantically). These formulations could be more or less detailed, but nevertheless contain the same intention. For example "gather information on products" vs. “checking advertising brochures” in the ‘household appliance script’, or “use brush and dishwashing liquid to rinse” vs. "clean dishes, glasses, cups and cutlery with hot water” in the ‘dish-washing script’. On the other hand, the same denotation can be intended to have a different meaning, dependent on the position of the action in the list. For example, “make inquiries about the possibilities for a trip” could either denote “checking friends’ options in terms of interests and dates”; or “find out possibilities once the actual destination of the trip is chosen (such as asking about prices and opening times of a museum or a restaurant)”. The judges retained the semantic meaning of the evocation in order to include an item in the frequency list. Elliptical answers were scored as an action only if they contained a verb and led to the stated goal. An ellipse is a descriptive or abridged remark that repeats the intended goal (e.g. “How would you get a new job?”, “I would apply for it”). A description of merely a location (e.g. for the sandcastle script, “on the beach”), or a material or tool (e.g. for the room painting script, “with a paint brush”) was not scored as an action.

Besides the above mentioned semantic aspects of the scripts there were also temporal aspects. Difficulties in the temporal aspect of the scripts were assessed by looking at sequencing errors, which correspond to a displacement in the logical sequence of actions within a script. Also, perseverative errors, which consist of actions that were repeated more than once in a script, were measured. With regard to open scripts obviously variations could occur in the order in which activities can be carried out without changing the end result. In these scripts we only scored physically impossible and illogical sequences as errors. A model script is provided in table 1.
Results

Data analysis
All data were analysed for normality first. When data were skewed either in
the patient sample or the control sample, non-parametric statistical tests
were used. Repeated measures ANOV A’s were used for multiple comparisons.
ANOV A’s are generally robust to moderate violations of normality
assumptions as long as sample sizes are not unreasonably small.

Neuropsychological tests
In table 2 the test performances and DEX-scores of patients and healthy
control subjects are reported. Mann-Whitney U tests for independent samples
showed significant differences between these groups on several neuro-
psychological measures (exceptions are TMT B/A, BADS Rule Shift, BADS
Action Plan, BADS Key Search, and BADS Temporal Judgement). When
significance levels were adjusted for multiple comparisons using the Holm-
Bonferroni method, all significant differences between patients and controls
were preserved.

Table 1  A model script: ‘purchase a household appliance, such as a washing
machine’

<table>
<thead>
<tr>
<th>Actions</th>
<th>Action Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explore options and gather information (advertising brochures, internet, consumers’ magazines, etc.)</td>
<td>major</td>
</tr>
<tr>
<td>Decide on the requirements of the appliance (such as capacity, energy label, size, etc.)</td>
<td>trivial</td>
</tr>
<tr>
<td>Find out on prices and decide on budget</td>
<td>major</td>
</tr>
<tr>
<td>Go to shop and inquire (about quality, prices and bargains)</td>
<td>minor</td>
</tr>
<tr>
<td>Compare the characteristics of several appliances (price, quality and requirements)</td>
<td>major</td>
</tr>
<tr>
<td>Choose and make a final decision</td>
<td>minor</td>
</tr>
<tr>
<td>Buy</td>
<td>major</td>
</tr>
<tr>
<td>Arrange to get the appliance home</td>
<td>trivial</td>
</tr>
</tbody>
</table>

Table 2  Performance in baseline neuropsychological tests: mean raw scores
(and standard deviations) and results of Mann-Whitney U tests
(one-tailed) of group differences for patients and healthy control
subjects

<table>
<thead>
<tr>
<th>Test</th>
<th>Patients (N = 48) Mean (SD)</th>
<th>Controls (N = 49) Mean (SD)</th>
<th>Z and p (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BADS Rule Shift</td>
<td>1.04 (1.79)</td>
<td>.96 (2.11)</td>
<td>z = -1.026; p = .153</td>
</tr>
<tr>
<td>BADS Action Plan</td>
<td>4.76 (7.1)</td>
<td>3.73 (8.4)</td>
<td>z = -4.29; p = .334</td>
</tr>
<tr>
<td>BADS Key Search</td>
<td>12.22 (3.83)</td>
<td>12.86 (3.27)</td>
<td>z = -8.88; p = .187</td>
</tr>
<tr>
<td>BADS Temporal Judgment</td>
<td>2.16 (8.5)</td>
<td>2.37 (8.8)</td>
<td>z = -1.249; p = .106</td>
</tr>
<tr>
<td>BADS Zoo Map</td>
<td>8.13 (4.79)</td>
<td>11.65 (3.80)</td>
<td>z = -3.538; p = .000</td>
</tr>
<tr>
<td>BADS Mod. Six Elements Test</td>
<td>5.25 (1.34)</td>
<td>5.84 (4.2)</td>
<td>z = -2.885; p = .002</td>
</tr>
<tr>
<td>BADS Total standard score</td>
<td>95.62 (12.64)</td>
<td>104.16 (10.15)</td>
<td>z = -3.402; p = .001</td>
</tr>
<tr>
<td>TMT B/A</td>
<td>2.21 (1.02)</td>
<td>2.34 (.81)</td>
<td>z = -1.151; p = .125</td>
</tr>
<tr>
<td>Stroop 3/2</td>
<td>1.62 (.26)</td>
<td>1.49 (.18)</td>
<td>z = 2.543; p = .006</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>36.21 (8.21)</td>
<td>44.07 (9.63)</td>
<td>z = -3.822; p = .000</td>
</tr>
<tr>
<td>Tower of London</td>
<td>10.67 (1.36)</td>
<td>11.51 (.74)</td>
<td>z = -3.902; p = .000</td>
</tr>
<tr>
<td>20 Questions Test</td>
<td>14.35 (4.67)</td>
<td>9.20 (4.65)</td>
<td>z = -4.845; p = .000</td>
</tr>
<tr>
<td>DEX self</td>
<td>31.00 (14.18)</td>
<td>18.11 (8.47)</td>
<td>z = -4.480; p = .000</td>
</tr>
<tr>
<td>DEX proxy</td>
<td>28.09 (13.34)</td>
<td>18.45 (9.92)</td>
<td>z = -3.374; p = .001</td>
</tr>
</tbody>
</table>

Note: BADS = Behavioural Assessment of the Dysexecutive Syndrome; TMT = Trail Making Test; VF = Verbal Fluency (animals and professions); TOL = Tower of London (total number correct); 20Q = Twenty Questions Test (number of questions asked); DEX = Dysexecutive Questionnaire.
Semantic aspects of scripts
Patients and healthy control subjects did not differ significantly form each other with respect to the total number of relevant actions ($t = -1.612, p = .109$). The proportion of irrelevant actions compared to the total number of actions (irrelevant / relevant + irrelevant) differed significantly between the two groups ($z = -3.144, p = .002$). Patients evoked significantly more irrelevant actions than control subjects ($z = -2.862, p = .004$).

An ANOVA using a repeated measures design was carried out with the four action types (major vs. minor vs. trivial vs. relevant intrusions), in which the assumption of sphericity was met. The results showed that there was a main effect for Action type ($F(1, 145) = 2.600, p = .109$). However, there was a Group x Action type interaction ($F(3,145) = 5.037, p = .002$). Tests of simple effects (4 One-way ANOVAs) showed that healthy controls evoked significantly more major steps ($F(1,145) = 4.233, p = .041$) and minor steps ($F(1,145) = 8.895, p = .003$) than patients. Patients tended to produce more trivial actions, but this difference was not significant ($F(1,145) = .851, p = .358$). The groups did not differ significantly with respect to relevant intrusions. These results show that the healthy controls produce more central actions (major and minor steps) than the patients.

Temporal aspects of scripts
Mann-Whitney U tests for independent samples showed that patients and healthy controls differed significantly in the number of perseverations ($z = -3.755, p = .000$), but not with respect to the number of sequencing errors ($z = -1.483, p = .138$). Patients made significantly more perseverative errors.

Script type
A repeated measures ANOVA by means of a 2 (patients vs. controls) x 2 (open vs. closed) x 2 (high vs. low frequency) was performed for relevant actions. This ANOVA showed main effects of structure ($F(1, 145) = 37.675, p = .000$) and frequency ($F(1, 145) = 76.443, p = .000$). Thus, open scripts or high-frequency scripts produced significantly more relevant actions than closed scripts and low-frequency scripts respectively. There was no significant group x structure interaction. Thus, both patients and control subjects produced more relevant steps in open than in closed scripts. There was a significant frequency x group interaction ($F(1, 145) = 15.364, p = .016$). One-way ANOVAs showed that there was a significant difference between the groups in low-frequency scripts ($F(1,145) = 6.914, p = .009$), but not in high frequency scripts. Healthy control subjects evoked significantly more relevant actions in the low-frequency scripts. Also, the group x structure x frequency interaction was significant ($F(1,145) = 5.489, p = .020$). Tests of simple effects showed that there was only a significant difference between both groups in the ‘open-low’ scripts ($F(1,145) = 7.075, p = .006$). Patients evoked significantly less relevant actions in the open-low-frequency scripts.

With respect to irrelevant actions, a 2 x 2 x 2 repeated measures ANOVA showed that there was a main effect of structure ($F(1, 145) = 12.079, p = .001$) and a main effect of frequency ($F(1, 145) = 6.942, p = .009$). Scripts with an open structure or low-frequency scripts evoked more irrelevant steps respectively. There was neither a significant group x structure interaction, nor a group x frequency interaction, nor a group x frequency x structure interaction. A similar result was found for sequencing errors: open scripts and low-frequency scripts yielded more sequencing errors. No significant interaction effects were found. Regarding perseverative errors a main effect was found for structure ($F(1,145) = 34.641, p = .000$) with a 2 x 2 x 2 repeated measures ANOVA, as well as a main effect for frequency ($F(1,145) = 18.568, p = .000$) and a group x frequency interaction effect ($F(1,145) = 10.631, p = .001$). Patients made considerably more perseverative errors than healthy control subjects in low-frequency scripts.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients (N = 48) Mean (SD)</th>
<th>Controls (N = 99) Mean (SD)</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>12.63 (2.875)</td>
<td>13.69 (2.961)</td>
<td>.016</td>
</tr>
<tr>
<td>Minor</td>
<td>5.71 (2.297)</td>
<td>6.98 (2.483)</td>
<td>.006</td>
</tr>
<tr>
<td>Trivial</td>
<td>5.10 (2.587)</td>
<td>4.73 (2.184)</td>
<td>.414</td>
</tr>
<tr>
<td>Relevant intrusions</td>
<td>2.31 (2.033)</td>
<td>2.35 (2.215)</td>
<td>.849</td>
</tr>
<tr>
<td>Total relevant actions</td>
<td>25.75 (7.312)</td>
<td>27.75 (6.911)</td>
<td>.109</td>
</tr>
<tr>
<td>Irrelevant intrusions</td>
<td>2.77 (3.328)</td>
<td>1.35 (1.751)</td>
<td>.004</td>
</tr>
<tr>
<td>Perseverations</td>
<td>5.00 (6.385)</td>
<td>1.92 (1.947)</td>
<td>.000</td>
</tr>
<tr>
<td>Sequencing errors</td>
<td>2.29 (1.650)</td>
<td>1.88 (1.500)</td>
<td>.380</td>
</tr>
</tbody>
</table>
Discussion

The aim of the present study was to assess script generation in a group of brain-injured patients with a dysexecutive syndrome. Previous research has shown that script generation is sensitive to frontal lobe damage. We investigated a mixed population of brain-injured patients with more diffuse lesions and lesions not only located in prefrontal regions. These patients had functional executive impairments and were included in the study on the basis of daily life executive complaints. Brain-damaged subjects with a dysexecutive syndrome are often referred for neuropsychological rehabilitation, and script tasks might not only substantiate that generating scripts involves a larger neural network than prefrontal tissue exclusively, but can also provide clues for the treatment of executive deficits in these specific patients. Problems with script generation might underlie patients’ complaints in the execution of daily life activities and therefore constitute a target element for treatment.

Our finding that patients and healthy control subjects evoked an equal number of relevant actions is in accordance with previous studies (Cazalis et al., 2001; Dritschel et al., 1998; Fortin et al., 2003; Godbout et al., 2005; Sirigu et al., 1995; Zanini et al., 2002). This suggests that the production of relevant action steps is not dependent on lesion location and appears not to be sensitive to brain damage. This implies that brain-injured patients with executive complaints are able to retrieve stored event knowledge. Only in one study (Godbout et al., 2004) it was found that both frontal lobe patients and parietal patients evoked less relevant actions than healthy control subjects.

However, the results of the present study showed that patients with a dysexecutive syndrome evoked significantly less central actions than control subjects, i.e., those actions that are the most essential in performing the activity (Allain et al., 2001). Hence, patients might be missing crucial elements in the planning of complex action sequences (see also Chevignard et al., 2000). These omissions can be interpreted as “goal neglect” (Duncan, Emslie, Williams, Johnson, & Freer, 1996), in which there is a mismatch between a stated goal and the goal list that needs to be generated to achieve this goal. Another interpretation of the lack of central actions in patients with dysexecutive problems is in terms of the importance of an action (Sirigu et al., 1995). Patients might have a problem with the judgment of the importance of actions i.e. the degree to which an action is perceived to lead to the stated goal. The fact that patients evoked less essential actions (major and minor), but did not evoke less accessory actions (trivial and relevant intrusions) confirms the results of Sirigu et al. (1995), who found that patients with frontal lesions made deviant estimates of action importance.

Moreover, the patients in our study evoked significantly more irrelevant actions compared to healthy control subjects (see also Fortin et al., 2003; Zanini et al., 2002). This also confirms the patients’ failure to focus on the essential actions of a script and their inability to inhibit irrelevant intruding thoughts.

Contrary to Fortin et al. (2003), Godbout et al. (2004), Godbout & Doyon (1995), and Sirigu et al. (1995), the patients in our study did not have problems with sequencing actions within scripts. This finding indicates that problems with script sequencing might be ascribed to purely frontal damage and thus might be a sensible indicator of prefrontal damage. On the other hand, patients in our study made considerably more perseverative errors than control subjects. This can be interpreted in terms of a problem with updating contents in working memory. Apparently holding script information online is difficult for patients with executive problems and comparable to the difficulties that have been reported in patients with prefrontal lesions (e.g. Godbout et al., 2004). Moreover, these perseverations were specifically found in low-frequency scripts, in which the element of novelty demands high degrees of executive control.

In fact, we think that low-frequency script generation tasks are very demanding in terms of cognitive control. Clearly, in order to produce these scripts, one has to maintain various information elements in a highly active state (the theme of the script, the set of relations between actors and actions in this script, the goal hierarchies, and the temporal order of events), and to suppress representations that might interfere with the to-be-remembered steps. One has also to actively hold in working memory the representation of the sub-goals of the script and to update them throughout the production of the different steps in order to avoid step repetitions. Finally, as in verbal fluency paradigms, it is likely that clustering and switching mechanisms are engaged to move from one action to another within a given scene (clustering) and from one scene to another scene within a given script (switching).

With regard to the structure of the scripts, open scripts and high-frequency scripts evoked significantly more relevant actions than closed scripts and low-frequency scripts respectively. Within open scripts more relevant actions can be retrieved from memory while describing how to perform activities. High-frequency scripts have a ready-scripted plan available which may enhance memories contained in a more specific database.

Healthy control subjects evoked significantly more relevant actions within the low-frequency category. This means that control subjects are able to exert more executive control in producing relevant actions in novel and non-routine tasks. Patients evoked significantly less relevant actions in the
open-low-frequency scripts. Similar to Dritschel et al. (1998), patients with executive problems were more impaired in scripts that require the highest degree of executive control, namely in open-ended, uncommon tasks.

Open scripts and low-frequency scripts also elicited the highest number of irrelevant actions. These types of scripts both call for the inhibition of automatic responses which do not necessarily lead the goal. Similarly, sequencing errors and perseverative errors occurred more in open scripts and low-frequency scripts. In short, there is considerable evidence that open scripts and low-frequency scripts are the most sensitive items in the script generation task that trigger dysexecutive symptoms, i.e. inefficient planning.

In summary, we found clear problems with script generation in a group of patients with more diffuse brain damage and lesions not restricted to prefrontal regions. Apparently, the generation of scripts requires an elaborate cortical network and is not exclusively controlled by the frontal lobes. This also means that not just patients with (pure) frontal lesions are in the need of treatment for planning problems. This study shows that the number of patients eligible for executive function training is considerable. Treatment should not only involve the planning of familiar tasks, but should comprise open-ended and novel problem-solving in particular. Current treatments for executive impairments, such as Goal Management Training (Levine, et al., 2000), can be complemented with planning exercises aimed at formulating intended activities and tasks in terms of goals and steps leading to these goals. Subsequently, patients should be taught how to put these action steps in the correct order.

This study has several limitations. Due to the absence of imaging data the conclusion that (aspects of) script generation rely on a large neural network mediates can only partially be substantiated. It is recommended that in future studies with a similar patient population damage localization data is provided. A second restraint concerns the lack of a second control group. Studies including patients with executive problems based on complaints and neuropsychological tests should include a control group of brain-damaged patients without executive problems to control for the nonspecific effects of brain damage (e.g. slow information processing, see Cazalis et al., 2001). A third recommendation for future research is about the execution of script tasks. In most studies concerning script generation, the successive execution of scripts is not performed. It could well be that patients who have difficulties in script information processing can execute these plans quite well, using environmental cues that trigger the concomitant actions. These cues are absent in the imaginary evocation of scripts. On the other hand, cues might also disturb the execution of scripts, by triggering irrelevant and impulsive actions (Chevignard et al., 2000; Duncan et al., 1996; Fortin et al., 2003).


Rehabilitation of executive disorders after brain injury: Are interventions effective?

Published in:
Journal of Neuropsychology 2011; 5, 73-113
D.H.E. Boelen, J.M. Spikman, L. Fasotti
Abstract

In this paper the effectiveness of interventions for executive disorders was reviewed. The objective was to evaluate the internal and external validity of intervention studies. A total of 46 papers, describing 54 studies, conducted in the last two decades meeting several preset inclusion criteria, was included in this review. The studies were categorised into three treatment approaches in order to enhance comparability. The overall results show that many interventions yield positive outcomes and seem to be effective in reducing executive problems in brain-injured subjects. However, several studies have only an explorative intent and are based on less sophisticated experimental designs. The verification of their results is generally more tenuous. The internal validity, or the set-up of experimental conditions necessary to draw valid conclusions about treatment effectiveness, including the choice of well-matched control groups, or the randomisation of patients over treatment and control conditions, is not always strong. The same conclusion can be drawn for the external validity of a number of the intervention studies; often evidence of generalization to real-life situations, long-term follow-up and transfer to non-trained situations, were (partially) lacking in the studies under review. The authors are aware that the design of proper randomized controlled trials (RCT) for the investigation of the treatment effectiveness of executive disorders is cumbersome and time-consuming. Nonetheless, the provisional results of several well-designed studies described in this review make the effort worthwhile.

Introduction

Many patients with acquired brain damage, following a stroke or traumatic brain injury, experience problems with planning, initiating, regulating and monitoring goal-directed behaviour (Baddeley & Wilson, 1988; Burgess and Robertson, 2002; Duncan, 1986; Lezak, 1995; Luria, 1973; Shallice, 1982). These problems are collectively described as executive dysfunction. The term executive dysfunction refers to an impairment of those cognitive processes and behaviours that enable a person to direct and control his or her behaviour in a purposeful and adequate way.

Traditionally, executive dysfunction was exclusively related to damage to the (prefrontal cortex, with frontal and executive functions considered as synonymous. More recently however, it has been shown that brain damage distant from the frontal lobes (e.g., Schmahmann & Sherman, 1998), damage to subcortical structures or interruption of connections between frontal and non-frontal areas, may also impair executive functions (e.g., Alvarez & Emory, 2006; Heyder, Suchan, & Daum, 2004; Miller & Cohen, 2001). Patients with executive problems may show an extensive array of symptoms. These may vary widely in degree and extent from patient to patient. Besides cognitive symptoms, behavioural, emotional and motivational disturbances may be observed, such as apathy, indifference, carelessness, impulsivity, irritability and disinhibition (e.g., Lezak, 1995; Mattson & Levin, 1990; Oddy, Worthington, & Francis, 2009). A complicating matter is the lack of awareness often shown by patients with frontal lesions, i.e., a deficient critical attitude toward their own impairment (e.g., Prigatano & Schacter, 1991).

Executive dysfunction may have a highly disrupting influence on patients’ lives. While independent functioning is possible to a certain extent with other cognitive disorders, such as memory or attention impairments, autonomous functioning may be directly compromised in the case of executive disorders (Lezak, 1995). Patients are unable to organize and monitor their behaviour properly, resulting in either passivity, or chaotic, impulsive behaviour. For example, they may be unable to complete daily pursuits, such as household chores, or fulfil vocational demands.

Given the impact on patients’ and their families’ lives and the social consequences of executive dysfunction, effective treatment is needed. The effectiveness of interventions for executive problems has been discussed by a number of reviewers, who have expressed their concerns about methodological constraints in the evaluation studies published so far (Cicerone et al., 2000, 2005; Kennedy et al., 2008; Turner & Levine, 2004). In the following review the evidential base for the effectiveness of interventions
targeting executive dysfunction will be investigated by focusing on two issues: the internal and external validity of intervention studies.

Internal validity is the degree to which valid inferences can be made about the effects of a treatment. Thus, internal validity concerns the soundness of an investigation. In order to allow for inferences with a high degree of internal validity, precautions must be taken during the design of the investigation. A study has poor internal validity if alternative explanations for the outcomes cannot be ruled out. With regard to external validity, issues of generalization and the relevance of results for daily life are important criteria for success. Generalization refers to applying treatment results to other tasks, settings and moments in time.

This paper reviews 46 papers covering 54 studies investigating the efficacy of executive function rehabilitation interventions for persons with acquired brain injury. We will critically review the papers in terms of their internal (i.e., experimental design) and external validity (i.e., generalisability of the impacts). Studies will be ranked according to their degree of validity. However, it should be noted that many of the reviewed studies are not necessarily flawed in this sense, because they only present preliminary evidence simply showing a treatment effect, without asserting any other claim. The studies are divided into three treatment categories: compensatory approach, which is further divided in internal and external strategies, restorative approach, and behaviour modification. We further stratify papers within each of these categories by level of evidence from randomized controlled trials to single case study designs.

Although many of the studies in the current review have been scrutinized before, this was mostly the case in non-peer reviewed journals and usually in a descriptive and non-systematic form (e.g., Turner & Levine, 2004). Also, these older reviews do not include the most recent literature. For example, the Ciccone et al. (2005) study reviewed papers that were published until 2002. Our study includes 16 papers published after 2002. As to the most recent review (Kennedy et al., 2008), which indeed is an analysis published in a peer-reviewed journal and with the most recent studies as a subject matter, it is not as comprehensive as our study. The Kennedy study reviewed merely interventions for executive functions after traumatic brain injury. Therefore, the Kennedy study focussed on a limited number of 15 studies, whereas our own study comprises the full range of rehabilitation papers on acquired brain damage of a non-progressive nature, totalling 54 studies. Moreover, the Kennedy study was limited to interventions for improving problem-solving, planning and organizing. We dealt with the complete myriad of symptoms that patients with executive problems may exhibit.

Method

The comprehensiveness of executive disorders and the variety of symptoms patients may exhibit have led to a diversity of treatment approaches. These approaches may aim at different agents for change and in this review studies were grouped into three general treatment approaches. Each approach targets a different aspect of executive impairment. The compensatory approach focuses on teaching patients new cognitive strategies to circumvent problems caused by executive impairments. The restorative approach specifically targets underlying executive impairments with the goal of restoring the deficit. Behavioural interventions seek to change disruptive behaviours caused by executive dyscontrol. Obviously, many more classifications are possible and the principles for classifying treatment studies are still open to debate (see, for example, Kennedy et al., 2008; but also Evans, 2005; Worthington, 2005). The internal validity of every intervention study within each of these categories was examined by judging the adequacy of research methods. The external validity was assessed by addressing the question if and how much the results generalized to situations beyond the treatment (environment) and after the treatment had ended.

Internal validity is the extent to which post-training changes can be causally attributed to the treatment. If there is a lack of internal validity, variables other than the independent(s) being studied may be responsible for part or all of the observed effect on the dependent variable(s). For example, selection bias refers to the presence of significant and systematic differences between groups before the start of treatment. Another example concerns the quality of outcome measures that were used to measure treatment effects, specifically with regard to test-retest effects. The available studies in this paper were ranked according to their methodological design and rigor. Although single-case designs certainly have their merits and are in many cases experimentally well designed, group studies are discussed first here, followed by the single-case studies, albeit they were discussed along the same lines. Within each treatment approach the studies with the highest degree of experimental control were listed first followed by studies with lower degrees of internal validity:

1) True experimental studies control for threats to internal validity by randomising the allocation of subjects to conditions. A true experiment is prospective; it is thought to provide convincing evidence of a causal relationship between variables. Randomized controlled trials (RCT) have become the ‘gold standard’ method for evaluating the efficacy of psychosocial interventions. (Nezu & Maguth-Nezu, 2008). Randomizing
participants makes it likely that the groups are comparable in all aspects except the nature of the intervention. This ensures that the ensuing outcomes can be solidly attributed to a given treatment.

2) Quasi-experimental group studies resemble an experiment in that they include both an experimental and a control group, but they lack the randomized allocation of trainees to these groups. Internal validity can be enhanced by matching participants beforehand on relevant characteristics, such as age, educational level, or intelligence.

3) Pre-experimental studies do not use a control group and therefore alternative explanations may have influenced the observed results.

Single-case studies are methodologically distinct from group designs (Tate et al., 2008) and were discussed separately within each of the three treatment approaches. They were also ordered following their degree of experimental control. For example, multiple baseline designs and single-case reversal designs with three or more phases (baseline or intervention, e.g., ABAB design or ABC design) have a higher degree of internal validity and therefore allow more convincing conclusions than single-subject designs involving one patient under only one non-repeated condition (e.g., AB designs or pre-test post-test designs). Single case studies have always been of great importance for neuropsychology. The role of single cases has been recognized as critical in the historical and theoretical development of the field (e.g., Caramazza, 1986). The main advantage of the single case design in (cognitive) rehabilitation is its capacity to tailor an intervention to the individual's characteristics. It therefore has an immediate application to clinical practice whereby its outcome provides empirical support for the intervention. However, a limitation of the single-subject design is its restricted ability to generalize beyond the individuals studied. Replication studies are required to ensure the same evidential value as group studies.

Although RCTs are reputed to provide the highest standards for establishing treatment effectiveness, we acknowledge that other experimental designs may also be particularly useful in the evaluation of interventions. In fact, RCTs are difficult to conduct, not only due to time and expense involved, but also because of ethical constraints (e.g., moral concerns when providing a no-treatment or placebo control condition). Often, studies with a less sophisticated design are conducted to provide preliminary evidence for an intervention, before conducting larger between-group comparisons. Such studies are used to examine novel treatment approaches.

External validity refers to the extent to which the results of an intervention study can be generalized beyond the conditions of the specific study to other individuals, tasks, settings, and over time. The relevance of results in the daily lives of patients is an important criterion of success. According to Worthington (2005) treatment for executive dysfunction should have an impact on behaviour in the real world. Therefore, three types of generalization of training effects were examined in this paper. Each study was assessed by considering three questions: Did the treatment effects apply to tasks other than the ones used in the treatment? Did the treatment effects transfer to the environment outside the treatment setting (i.e., what is the relevance for daily life)? And, were follow-up measures taken to ensure that change was maintained over time? As to the different treatment approaches advocated in this study, generalization might be expected to occur in different forms. In the compensatory approach, for example, patients are often taught new, general ways to compensate for problems in everyday life resulting from cognitive impairments. Thus, the occurrence of generalization in the form of spontaneous application of strategies to untrained tasks is expected. In fact, such an intervention is not aimed at relearning specific tasks, but at teaching new ways to handle problems, despite the lasting presence of the cognitive deficits (Geusgens, Winkens, van Heugten, Jolles, & van den Heuvel, 2007). It, however, the strategy has been developed for the relearning of a specific task, generalization to untrained tasks can not be expected. Restorative interventions, on the other hand, aim to improve particular cognitive processes in a broad sense. Therefore, if such an intervention is effective, this should be reflected in the generalization to untrained tasks in which the restored process is needed (e.g., Coltheart, Brunsdon, & Nickels, 2005). Finally, in the category of behaviour modification studies, transfer of treatment effects is dependent on the techniques used. The gains of some techniques may be limited to the environment where the intervention took place, whereas in other techniques progress may generalize to other environments (see Alderman, Fry, & Youngson, 1995; Knight, Rutterford, Alderman, & Swan, 2002).

A PsycINFO and MEDLINE literature search was conducted using a combination of the following search words: (cognitive) rehabilitation, brain damage, (traumatic) brain injury, stroke, frontal lobe damage, frontal lobe dysfunction, dysexecutive syndrome, executive functions, executive disorders, executive deficits, executive dysfunction, organizational deficits, executive impairments, retraining, treatment, problem-solving, remediation, self-monitoring, planning disorder, executive control, behaviour dyscontrol, and behaviour management. Also, reference lists from the identified papers were searched to increase the initial list of references. In addition, the internet site ‘PsycBITE’ (http://www.psycbite.com) which is an interactive database of interventions for the psychological consequences of acquired brain
improvement in 30 traumatic brain injury (TBI) patients. GMT was designed to facilitate the maintenance of intentions in goal-directed behaviour (Robertson, 1996). Patients were randomly assigned to either GMT or Motor Skills Training (MST). Training effects were measured by three paper-and-pencil simulations of real-life tasks. Results show that GMT was associated with improved performance, i.e., slowing down on post-training tasks, indicating increased care and attention to the tasks, and a reduction in errors. However, as two of the tasks that were incorporated in the training were also used as post-training measures, they mainly may have measured practice effect. Still, the GMT group showed a significant reduction in errors on the third, untrained task. Other (everyday life) generalization tasks were not employed. Follow-up assessment was not carried out.

Von Cramon, Matthes-von Cramon, and Mai (1991) conducted a randomized group study in which a step by step method was introduced to analyse complex, multistage problems. Patients with various aetiologies of brain damage and with documented problem-solving impairments, nearly all of whom were at a chronic stage of their brain injury, were allocated to either the Problem-Solving Training (PST) or Memory Training (MT). Treatment effects were evaluated by using neuropsychological problem-solving tests and intelligence tests. Also, an ecological planning test and a rating scale for everyday problem-solving behaviour, as judged by staff members, were used to measure generalization of treatment effects. The trained tasks did not resemble the pre-post tests. Results showed improvement on problem-solving tests, planning tests (including the ecological planning test) and intelligence tests, in favour of the PST group. The results suggest that patients in the PST group were able to learn a slower, more controlled processing of given problems. The PST group also improved on the behavioural rating scale, which suggests that some patients were able to transfer problem-solving skills to everyday situations on the ward. However, maintenance of skills outside the ward was not investigated. Follow-up measurements did not take place.

An elaboration of problem-solving training was presented by Rath, Simon, Langenbahn, Sherr, and Diller (2003), who treated 46 people with TBI in a randomized controlled study. Strategies addressing emotional self-regulation were incorporated, which was held to be a prerequisite to successful (social) problem solving. Patients were randomly assigned to either ‘conventional neuropsychological group treatment’ or the innovative problem-solving treatment. A variety of measurements was used, i.e., neuropsychological tests of attention, memory, reasoning and executive function, ratings of community integration, symptom complaints, and self-esteem.
measures could not be rated blindly. The internal validity of this study was
in the control group on these four measures. Unfortunately, one of the effect
experimental group improved significantly more over time than the patients
specifically targeted daily life executive functioning. Patients in the
treatment. Both the treatment itself as well as four outcome measures
Assessment took place before, immediately after, and 6 months post-
(ABI) of various aetiologies, all of whom were in the chronic stage of recovery.
function training. Participants were 75 patients with acquired brain injury
remediation, self-inhibition, flexibility, and strategic behaviour. The treatment
framework: self-awareness, goal setting, planning, self-initiation, self-moni-
toring, self-inhibition, flexibility, and strategic behaviour. The treatment
relied on Goal Management Training (Levine et al., 2000) and Problem
Solving Training (von Cramon et al., 1991; von Cramon & Matthes-von
Cramon, 1994). It also involved a systematic Goal Planning Approach. This
treatment was compared with a computerized cognitive function training. Participants were 75 patients with acquired brain injury
(ABI) of various aetiologies, all of whom were in the chronic stage of recovery.
Assessment took place before, immediately after, and 6 months post-
treatment. Both the treatment itself as well as four outcome measures
specifically targeted daily life executive functioning. Patients in the
experimental group improved significantly more over time than the patients
in the control group on these four measures. Unfortunately, one of the effect
measures could not be rated blindly. The internal validity of this study was
mitigated by the fact that not each of the eight targeted executive aspects
were explicitly assessed.

Hewitt, Evans and Dritschel (2006) examined whether a self-instructional
technique involving self-cueing to recall specific autobiographical
memories would improve performance on a planning task (Everyday
Descriptions Task, EDT; Dritschel, Kogan, Burton, Burton, & Goddard, 1998).
Thirty subjects with TBI were randomly allocated to an experimental or a
no-treatment group. In the first group subjects received a training aimed to
prompt the retrieval of specific memories to support planning. The effects of
this training were measured by means of scores on the verbal reports of how
subjects would plan to solve several activities. Compared to the no-treatment
group the experimental group showed an increase in the number of specific
memories recalled and devised more effective plans with more relevant
steps. However, the study did not provide a follow-up assessment. Also,
without a placebo control group, one can not discount the possibility of a
non-specific effect of the training (e.g., effort). It was not investigated whether
the increased performance on the EDT was related to functional aspects of
everyday problem-solving.

Webb and Glueckauf (1994) randomly assigned 16 chronic patients with
TBI to either a High Involvement goal-setting group (who received explicit
strategies for prioritization and goal monitoring, using a diary and
worksheets), or a Low Involvement goal-setting group (receiving pre-assigned
goal lists, but no formal monitoring training). The aim of the study was to
assess the effects of direct involvement in goal-setting on rehabilitation
outcome. Both groups improved on ratings of goal attainment from pre-
to post-testing. Maintenance of gains in goal attainment at follow-up after two
months was restricted to the High Involvement group. The authors proposed
that the active self-monitoring and self-control aspect of the Goal Attainment
Scaling technique promoted this maintenance. There was no report on
generalization of the effects to other domains of daily life.

Pre-experimental group studies.
The effects of Interactive Strategy Modelling Training (ISMT) on verbal
problem-solving were examined by Marshall et al. (2004) in a preliminary study.
Twenty patients with TBI were taught question-asking strategies to enhance
their performance on solving problems that involved verbal reasoning.
However, these strategies were not described. After treatment patients asked
more strategic questions. Follow-up measurements 1 month after treatment
showed that this improvement was maintained. However, there was no
control group or control condition with which to compare the verbal
reasoning skills of the participants. Also, the trained task was similar to the pre-post tests. No information was given on generalization of training effects to other measures or to daily life situations.

Single-case studies.
Cicerone and Giacino (1992) sought to improve self-regulation in TBI patients, described three case studies. In the first of these studies two patients were taught to predict their performance on the Tower of London task and to adjust their efforts to the task demands. Results show that response latencies increased significantly, suggesting that they spent more time on planning the next steps in solving a problem. There was also a marked decrease in the number of errors in both patients. Brief anecdotal information is given on generalization of training effects. A multiple baseline across subjects design was used in the second part of the study, which concerned self-instructional training with verbal mediation in six patients. The training was delivered in three stages to promote the progressive internalization of verbal self-regulation, using the Tower of London task as a training task. Five out of six patients showed a marked reduction of errors on this trained task, representing a task-specific training effect. Evaluation of pre- and post training changes indicated significant differences on two psychometric measures. Anecdotal information suggests that behavioural regulation was enhanced in real-life behaviour in two patients who received additional generalization training. In the third study, an error-monitoring procedure was introduced using a reversal design to increase error recognition and self-monitoring in one patient. The procedure was effective in improving performance on the Tower of London task, although withdrawal of the treatment resulted in a return to the baseline error rate. These might also be ascribed to task-specific practice. Only with direct instruction was this patient able to apply the error-monitoring strategy to a real-life task. Follow-up data were not presented in the study.

Cicerone (2002) aimed to improve the ability of patients to effectively allocate attentional resources and to manage the rate of information during performance on computerized N-back tasks. Patients were taught compensatory strategies, such as self-pacing, verbal mediation, self-monitoring of their effort during performance and the use of feedback by therapists. This preliminary study employed a prospective case-comparison design with a sample of eight patients, four of which received treatment. Allocation to treatment was not random. Although the study was presented as a case-comparison design, the data were analysed at a group level. The mixture of single-case and group statistical methods does not convincingly support the conclusion that the intervention was effective. The results indicated that the intervention had a positive effect on tests of attention and working memory, with no direct effect on processing speed. Generalization was assessed by means of a questionnaire that showed a significant reduction of self-reported attentional difficulties in everyday activities in the treatment group. There were no follow-up measurements.

Von Cramon and Matthes-von Cramon (1994) presented the case of a medical doctor who 9 years after TBI received Problem Solving Training. The training procedure was embedded in a protected work trial, so that treatment in self-regulation could be studied under real-life conditions. The PST-steps were taught using self-instruction techniques. Goal Attainment Scaling was used to evaluate training effects on work performance. Also, a problem-solving simulation was administered before and after the training, as well as an awareness rating scale (see table 1a). Work performance improved, but training effects did not generalize to novel situations. The patient’s awareness did not increase and confronted with novel or changing situations his basic incompetence seemed to be still present. The study was not systematically followed up. However, the authors mention that the patient’s work reports were more coherent and that he was able to get hold of supported employment after the training period.

Goal Management Training (GMT) was used to instruct a postencephalitic patient how to prepare a meal (Levine et al., 2000). The outcome measures consisted of two everyday paper-and-pencil tasks, meal preparation performance (through observation), and the patients’ self-reports of meal preparation behaviour. Both observation and self-report measures revealed improved meal preparation performance after GMT, which was maintained in follow-up measurements at 1, 3, and 6 months. Apart from anecdotal information no other data on generalization to other situations and tasks were reported.

Schweizer et al. (2008) applied GMT to a patient with executive dysfunction following a right cerebellar hemisphere arteriovenous malformation haemorrhage. The effectiveness of GMT was assessed using a battery of standardized and experimental tests of executive function and attention and two questionnaires (see table 1a). The results show that the patient’s awareness increased and that his performance on tests improved, both immediately after rehabilitation and at follow-up after 48 days and 4 months.

Satish, Streufert, and Elingr (2008) applied strategic management simulation (SMS), initially designed to measure decision making competence in healthy adults within industrial-organizational settings, to assess the executive impairments of a male who had sustained a head injury and who
performed within normal limits on traditional neuropsychological tests. Following assessment, treatment of the patient focused on initiative, information utilization, response flexibility, and decision making, using decision-making and problem-solving vignettes of everyday life and vocational situations. A second, parallel scenario of SMS assessment showed that performance on the trained competencies had improved. The study did not report follow-up data, nor was generalization measured (although anecdotal information was provided).

Conclusion.

Sixteen intervention studies were discussed in the compensatory approach with internal strategies, six of which were true experimental group studies. Although the latter warrant firm conclusions about the effectiveness of the proposed treatments, several aspects of these studies may be considered. Every study was aimed at a different aspect of executive dysfunction (e.g., goal-management, problem-solving, goal-setting, time-pressure management), which renders the replication of studies difficult. In the majority of these studies one or more aspects of generalization were assessed. The studies of Rath et al. (2003) and Spikman et al. (2010) stand out as well-conducted RCTs with positive outcomes on real-life training tasks and with maintenance of effects. One pre-experimental study was identified. In the absence of a control group, valid conclusions about the effectiveness of the treatment might be more limited. With regard to the single-case studies, threats to internal validity were controlled by employing a multiple baseline across subjects design in one study (Cicerone & Giacino, 1992). In another single-case study internal validity was enhanced by replication over subjects (Cicerone, 2002) and in a third study by replication of treatment phases (Cicerone & Giacino, 1992). In the remaining five single-case studies the observed behavioural changes cannot be attributed to the treatment with certainty. Most of the single-case studies did not provide follow-up assessment, so it remains unclear whether treatment effects were maintained over time. All of the single-case studies reported on generalization, although in all but one study (Von Cramon & Mathes-von Cramon, 1994) this information was anecdotal. Another asset of the Von Cramon study was that problem-solving assignments were included to facilitate daily problem-solving. The results of the training were identical in each delivery mode. The subjects were given instructions to draw analogies to solve similar, new problems. Home assignments were included to facilitate daily problem-solving. The results showed increased self-efficacy and improved basic problem-solving skills in all three delivery modes. In a subsequent clinical trial (Man, Soong, Tam, & Hui-Chan, 2006), the same three conditions of problem-solving training were compared and a no-treatment control group was added. The results of 103 subjects who were randomly assigned to one of the four groups showed that the analogy-based problem-solving training was effective for improving

Compensation: External strategies

A total of 20 intervention studies described in 16 papers were classified in this category, of which 3 were RCTs, 3 were quasi-experimental group studies, 3 pre-experimental group studies, and 11 were single-case studies.

True experimental group studies.

Wilson, Emslie, Quirk, and Evans (2001) examined the effectiveness of NeuroPage, a portable pager system linked to a central computer, which sends reminders to its users. A RCT was used to assess achievement of everyday targets in 143 participants with memory, planning, attention and organization problems. A crossover design ensured that some participants received the pager after a 2-week baseline whereas others had to wait for 7 weeks. Compared to the baseline period, 121 out of 143 benefited significantly from the use of NeuroPage. Post-treatment data were available for 74 participants, of which 54 were still significantly more successful than at baseline, 7 weeks after returning their pager. From this success rate, it can be concluded that generalization took place to everyday life without a pager. The results of the subjects with TBI (N=63) were separately reanalysed, showing that for the majority of the TBI patients (81%) in this study the pager led to a significant increase in levels of independence (Wilson, Emslie, Quirk, Evans, & Watson, 2005). Similar results were found in four encephalitic patients who were significantly more successful at carrying out everyday tasks when using the pagers. (Emslie, Wilson, Quirk, Evans, and Watson, 2007). Finally, Fish, Manly, Emslie, Evans, and Wilson (2008a) reported similar findings for a subgroup of stroke patients (N=36).

Soong, Tam, Man, and Hui-Chan (2005) investigated which of three delivery modes of problem-solving training was most effective in a pilot study with 15 brain-injured subjects. These were randomly allocated to a computer-assisted skills-training programme, an online interactive skills training programme, or a therapist-administered programme. The contents of the training were identical in each delivery mode. The subjects were given source problems that they would encounter in everyday life and were instructed to draw analogies to solve similar, new problems. Home assignments were included to facilitate daily problem-solving. The results showed increased self-efficacy and improved basic problem-solving skills in all three delivery modes. In a subsequent clinical trial (Man, Soong, Tam, & Hui-Chan, 2006), the same three conditions of problem-solving training were compared and a no-treatment control group was added. The results of 103 subjects who were randomly assigned to one of the four groups showed that the analogy-based problem-solving training was effective for improving
<table>
<thead>
<tr>
<th>Study design</th>
<th>Participants</th>
<th>Time since injury</th>
<th>Control group</th>
<th>Intervention</th>
<th>Trained tasks</th>
<th>Outcome measures</th>
<th>Results</th>
<th>Follow-up assessment</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levine et al. (2000) -1</td>
<td>Randomized control design</td>
<td>TBI (30)</td>
<td>M=3.7 years</td>
<td>Patient control group (15): Motor Skills Training</td>
<td>Goal Management Training (GMT): stepwise technique to prevent goal neglect, support goal directed behaviour and self-regulation</td>
<td>Paper-and-pencil simulations of real-life tasks</td>
<td>Paper-and-pencil simulations of real-life tasks</td>
<td>GMT group improved compared to control group on paper-and-pencil simulations of real-life tasks</td>
<td>No report</td>
</tr>
<tr>
<td>Von Cramon et al. (1991)</td>
<td>Randomized control design</td>
<td>CHI, CVA, TBI, other (37)</td>
<td>Range 2-120 months (median=7)</td>
<td>Patient control group (17): Memory Training</td>
<td>Problem Solving Training (PST): stepwise technique to reduce complexity of multistage problems</td>
<td>Generation of goal-directed ideas, discriminating relevant and irrelevant information, processing multiple information</td>
<td>Intelligence tests, TOH, MCST, ecological planning test; rating of problem solving behaviour (observed by therapists)</td>
<td>PST group improved on 3/5 IQ subtests, TOH, ecological planning test, behavioural ratings of awareness, goal direction, problem solving and action style</td>
<td>No report</td>
</tr>
<tr>
<td>Fasotti et al. (2000)</td>
<td>Randomized control design</td>
<td>CHI (22)</td>
<td>Range 3-43 months (M=9.8)</td>
<td>Patient control group (10): non-specific concentration training</td>
<td>Time Pressure Management (TPM): stepwise procedure to compensate for executive problems brought about by slow information processing.</td>
<td>Nine videotaped short stories</td>
<td>Reproduction scores, number of preventive steps, number of managing steps of two videotaped short stories, behavioural observation list; tests of memory, attention, speed of processing; questionnaires of psychosocial well-being</td>
<td>TPM group: significant more managing steps and improved performance on tests of attention and memory; Both groups: improvement on information intake, no effect on prevention strategies, no improvement in well-being</td>
<td>Number of managing steps stable at 6 months; no improvement in well-being in both groups</td>
</tr>
<tr>
<td>Spikman et al. (2010)</td>
<td>Randomized control design</td>
<td>ABI (CVA, TBI, Tumour, other; 79)</td>
<td>Range 4-468 months (M=59.6; median =22)</td>
<td>Patient control group (37): computerized cognitive function training</td>
<td>Multiaxied treatment comprising Goal Planning Approach, Goal Management Training and Problem Solving Training</td>
<td>Didactic material, psycho-education, strengths and weaknesses analysis, roleplays and discussion of real-life problems reported by participants, home assignments</td>
<td>Role Resumption List (social participation), Treatment Goal Attainment, Executive Secretarial Task, DEX, EOS, QOLIBRI, BADS, TMT, Stroop Test, TOL, 19 words test</td>
<td>Experimental group improved significantly on outcome measures (RRL, TGA, EOS). Experimental group performed better on outcome measures (RRL, TGA, EOS, EST)</td>
<td>Improvement on measures of daily life (RRL, TGA, EOS) and also on ecological executive test (EST, i.e. problem solving simulation)</td>
</tr>
<tr>
<td>Study design</td>
<td>Participants</td>
<td>Time since injury</td>
<td>Control group</td>
<td>Intervention</td>
<td>Trained tasks</td>
<td>Outcome measures</td>
<td>Results</td>
<td>Follow-up assessment</td>
<td>Generalization</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------</td>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Hewitt et al. (2006)</td>
<td>Randomized control design</td>
<td>TBI (30)</td>
<td>At least one year</td>
<td>Patient control group (15): no treatment</td>
<td>Episodic memory cueing procedure: prompting the retrieval of specific memories to support planning</td>
<td>Various questions similar to EDT (verbal reports of how subjects would plan to solve activities)</td>
<td>Performance on EDT (effectiveness of the plan, number of steps, number of specific memories included)</td>
<td>Increase in number of specific memories recalled; increase in effectiveness of plan and number of relevant steps in the plan</td>
<td>No report</td>
</tr>
<tr>
<td>Webb &amp; Glueckauf (1994)</td>
<td>Randomized control design</td>
<td>TBI (16)</td>
<td>M=8.7 years</td>
<td>Patient control group (8): low involvement</td>
<td>Goal-setting training; formalized structuring of rehabilitation goals (orientation, goal-setting, monitoring)</td>
<td>Rehabilitation activities</td>
<td>Goal Attainment Scaling; diary (perceived goal progress, rate goal attainment)</td>
<td>Significant more progress in attaining goals in High Involvement group</td>
<td>Significant gains in goal performance from pre-testing to 2-months follow-up</td>
</tr>
<tr>
<td>Cicerone &amp; Giacino (1992) -2</td>
<td>Single-case design (multiple baseline across subjects)</td>
<td>TBI (5), tumour (1)</td>
<td>At least one year</td>
<td>Self instruction: internalization of verbal self-regulation</td>
<td>Tower of London (TOL)</td>
<td>TOL performance during training; neuropsychological tests; informal observation in real-life situations</td>
<td>Reduction of errors during monitoring procedure; reduction of problem behaviours</td>
<td>No report</td>
<td>Spontaneous application of strategies in untrained situations in 2 patients after additional generalization training</td>
</tr>
<tr>
<td>Cicerone &amp; Giacino (1992) -3</td>
<td>Single-case design (A-B-A-B)</td>
<td>TBI (2)</td>
<td>No report</td>
<td>Error monitoring, feedback, correction and modelling (guided rehearsal) of desired responses</td>
<td>Tower of London; socially inappropriate behaviour</td>
<td>TOL performance; number of problem behaviours</td>
<td>Reduction of errors during monitoring procedure; reduction of problem behaviours</td>
<td>No report</td>
<td>1st patient applied error monitoring routine in clerical task; recognition and self-correction of problem behaviours in 2nd patient</td>
</tr>
<tr>
<td>Cicerone (2002)</td>
<td>Prospective case-comparison design</td>
<td>Mild TBI (8)</td>
<td>M=8.25 months</td>
<td>4 control subjects (non-randomised): no treatment</td>
<td>Deliberate use of strategies to allocate attentional resources and manage rate of information during task performance</td>
<td>N-back task, N-back task with verbal generation, N-back dual-task</td>
<td>Tests of attention (TMT, PASAT, CPT, 2&amp;7 test); self-report questionnaires</td>
<td>3/4 experimental participants improved on CPT, PASAT, 2&amp;7 test; reduction of attentional complaints on questionnaires</td>
<td>No report</td>
</tr>
<tr>
<td>Study design</td>
<td>Participants</td>
<td>Time since injury</td>
<td>Control group</td>
<td>Intervention</td>
<td>Trained tasks</td>
<td>Outcome measures</td>
<td>Results</td>
<td>Follow-up assessment</td>
<td>Generalization</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>---------</td>
<td>---------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Von Cramon &amp; Matthes-von Cramon (1994)</td>
<td>Single-case design (AB)</td>
<td>TBI (1)</td>
<td>9 years</td>
<td>Problem Solving Training (PST): stepwise technique to reduce complexity of multistage problems</td>
<td>Vocational tasks</td>
<td>Vocational performance, Goal Attainment Scaling; computerized problem-solving simulation, awareness rating (PCRS)</td>
<td>Work performance improved; no change in problem-solving simulation or awareness rating</td>
<td>No report</td>
<td>No generalization of training to novel situations (i.e. problem-solving simulation and PCRS). Improvement of work performance. Maintenance of supported employment after 2 years</td>
</tr>
<tr>
<td>Levine et al. (2000) - 2</td>
<td>Single-case design (A-B)</td>
<td>Meningo-encephalitis (1)</td>
<td>5 months</td>
<td>Goal Management Training (GMT)</td>
<td>Meal preparation</td>
<td>Paper-and-pencil simulations of real-life tasks; observation of meal preparation; self-report diary</td>
<td>Improvement on paper-and-pencil tasks, meal preparation errors reduced (formal observation and self-report)</td>
<td>Gains stable at 1, 3, 6 months (meal preparation observation)</td>
<td>Anecdotal report of generalization to other activities at home and at work</td>
</tr>
<tr>
<td>Schweizer et al. (2008)</td>
<td>Single-case design (A-B)</td>
<td>Cerebellar AVM rupture (1)</td>
<td>127 days</td>
<td>Goal Management Training (GMT)</td>
<td>No report</td>
<td>Standardized and experimental tests of executive function and attention; DEX, CFQ</td>
<td>Increased awareness (CFQ, DEX); reduction in errors on SART and Tower test</td>
<td>Improved performance on CVLT, TMT at 48 days; gains stable on SART and Tower test at 4 months</td>
<td>Anecdotal information: patient returned to work after rehabilitation</td>
</tr>
</tbody>
</table>

Note. TBI, traumatic brain injury; CVA, cerebrovascular accident; CHI, closed head injury; AVM, arteriovenous malformation; TONI, Tower of Hanoi test; WCST, Wisconsin Card Sorting Test; EOS, Executive Observation Scale; QOLIBRI, Quality of Life after Brain Injury; BADS, Behavioural Assessment of the Dysexecutive Syndrome; EDT, Everyday Descriptions Task; CFQ, Cognitive Failures Questionnaire; PASAT, Paced Auditory Serial Addition Test; RAPS, Rapid Assessment of Problem Solving; CPT, Continuous Performance Test; ARMS, Attention Rating and Monitoring Scale; TOL, Tower of London test; WISC, Wechsler Intelligence Scale for Children; PCRS, Patient Competency Rating Scale; TMT, Trail Making Test; DEX, Dysexecutive Questionnaire; SART, Sustained Attention to Response Task; CVLT, California Verbal Learning Test
problem-solving skills, regardless of its delivery mode. Both in the pilot study (Soong et al., 2005) as well as in the clinical trial, care was taken to measure generalization of effects by the administration of a self-efficacy checklist (see table 1b). Maintenance of effects was not measured in either study.

Quasi-experimental group studies.
Fasotti, Bremer, and Eling (1992) addressed arithmetical word problem-solving in order to examine whether a cueing procedure aimed at the improvement of text encoding skills could ameliorate arithmetical word problem-solving. Arithmetical word problem-solving is a typical example of verbal reasoning, which is a crucial daily life executive skill. Thirty patients with frontal lobe damage, 10 patients with left posterior hemisphere lesions, and 10 healthy controls participated in the study. Both the patient groups and the healthy control group received treatment in solving cued and uncued verbal problems. Problem solving performance improved with systematic cueing in the left and bilateral frontal group, but not in the right frontal or the left posterior group. The effect of cueing on solution time was only significant in the bilateral frontal group. Maintenance of effects was not assessed, nor was generalization investigated. Although this was not a rehabilitation study in the strict sense of the term, investigating the effects of cueing in patients with frontal lobe damage can contribute to the development of cueing methods in other (clinical) studies (e.g., Delazer, Bodner, & Benke, 1998).

In a study by Manly, Hawkins, Evans, Woldt, and Robertson (2002), involving 10 brain-injured patients and 24 healthy controls, patients were exposed to periodic, non-predictive auditory alerts, while performing a 6-elements planning task (Hotel Task), in an effort to overcome goal neglect (Duncan, Emslie, Williams, Johnson, & Freer, 1996). The patients’ performance improved significantly in the cued condition, i.e., the number of tasks attempted and the time allocation to those tasks. In order to control for non-equivalence of the experimental and control group randomized crossed parallel versions of the planning task were used. There was no follow-up assessment and generalization of alerting effects was not investigated.

The effect of cueing on the maintenance of attentive control was further investigated in three successive experiments using the Sustained Attention to Response Test (SART), a go-no-go computerised task designed to measure absentminded lapses in action (Manly et al., 2004). The results of the first experiment showed that patients with a right hemisphere stroke performed worse on the SART than healthy control subjects. Both patients and controls benefited from the auditory cues. The effect of cueing was found to be independent of the subject’s current attentive state and to be the greatest in the predictable fixed-sequence condition of the SART, where responses were more automatic. The second and third experiment investigated cueing effects on healthy controls only and were therefore classified as pre-experimental studies. The study did not report on generalization of alerting effects and was not followed up. Similar to Fasotti et al. (1992), these experimental manipulations (Manly et al., 2002, 2004) were not strictly intended as rehabilitation studies. However, they were considered valuable in this paper because investigations into the effects of cueing in patients with executive problems can lay the foundations for more clinically oriented studies.

Pre-experimental group studies.
Fish et al. (2007) investigated the effects of cueing on prospective memory performance in a single group of patients under two conditions. In the cued condition text messages were sent to patients’ mobile telephones on five randomly selected days. The cues were ‘content-free’, that is unlike the NeuroPage system, they did not directly cue execution of a specific prospective memory task. Rather, they were designed to cue patients to engage in general review of their goals. The possibility of carry-over effects was controlled by randomizing the two conditions separately for each participant. Participants were required to make telephone calls to a voicemail service at four set times each day for ten days. There was a significant effect of cueing on the proportion of target telephone calls and on timing accuracy of these telephone calls. There was no follow-up assessment. An asset of this study was that the effects of cueing were examined in patients’ daily lives instead of in the clinic. However, transfer of cueing effects to other, untrained tasks was not investigated.

Single-case studies.
Wilson, Evans, Emslie, and Malinek (1997) conducted a pilot study to investigate the effects of NeuroPage. Fifteen neurologically impaired subjects with everyday memory and planning problems were included, to each of which a single-case reversal design was applied. It was demonstrated that all subjects were significantly more successful in completing everyday tasks with the pager than without. For some of the patients there was no decline in performance after the pager was withdrawn. Follow-up assessment was not carried out.

An ABAB single-case design was used to evaluate the impact of NeuroPage in a patient who was unable to initiate certain actions without being prompted (Evans, Emslie, & Wilson, 1998). Compared with the baseline phases, the
patient was more successful in completing a number of everyday tasks during each of the experimental phases with the pager. The use of a checklist shortened the time needed to complete a morning hygiene sequence. Transfer of the effect of the pager to untrained tasks was not investigated.

In a follow-up study (Fish, Manly, & Wilson, 2008b) the use of NeuroPage and a checklist was reintroduced in the same patient, who no longer used the pager 10 years after the initial intervention (Evans et al., 1998). Instead she had used a self-programmed electronic organizer, but without the apparent success of the pager. One possibility was therefore that the patient showed better compliance with cues that were perceived as coming from an external source. Assessment showed no deterioration in neuropsychological functioning. Whereas in the previous study the two strategies (pager & checklist) were introduced simultaneously and their effects examined upon separate goals, in the current study the two strategies were reintroduced separately, and their effects examined on three common goals. This was accomplished by applying an alternating treatments design. Visual analysis of the time-series data, as well as statistical analyses showed that performance during the NeuroPage-phase was significantly better than the checklist-phase for all measures. Information about a follow-up after 6 months was anecdotal.

Wade and Troy (2001) evaluated the use of mobile phones to provide memory prompts for individuals with brain injury. Five chronic head injury patients with memory and executive impairments were selected. Self-initiated performance in remembering to carry out target tasks were counted both with and without a mobile phone. Baseline measurements consisted of neuropsychological tests, while the counting of self-initiated target activities recorded in individualized diaries took place both at baseline and during the treatment. The use of mobile phones appeared to be effective for all five users, although the experiment was discontinued in one participant. Another participant did not use the phone consistently and the caregiver compliance with the diary proved variable, so that the data were incomplete. Follow-up assessment did not take place. The use of simple AB-designs makes it difficult to attribute treatment effects to the use of the mobile phone as a memory aid.

Kirsch, Levine, Lajiness-O’Neill, and Schnyder (1992) investigated the use of interactive task guidance (ITG), a computerized intervention that is intended to facilitate the performance of functional tasks through the use of cues that were presented to patients on a computer screen. Four chronic patients with TBI participated in the study. A single-case reversal design was used for each participant, comparing the performance on a vocational task, using either written directions, or ITG. In both conditions, the steps necessary to complete the task were presented, although in the ITG condition each direction was followed by the instruction to “press the letter C to continue” when the current step was completed. ITG thus monitored the subject’s performance and offered corrective feedback. Two of the patients clearly benefited from ITG, while the other two presented more equivocal responses. Baseline levels were established in a trial with written directions that was also used as a control condition. This may have diminished contrast between the treatment conditions. Therefore, it remains unclear how the subjects would perform without cueing.

Burke, Zencius, Wesolowski, and Doubleday (1991) presented six case studies in which the aim was to improve several aspects of executive functioning. The first case described a patient who was taught to use task-based checklists to enhance planning and problem-solving in four vocational tasks. A multiple baseline design across behaviours was used, which allows to conclude that the checklist indeed improved the ability of the patient to complete the tasks autonomously. Performance improved during the fourth baseline without the introduction of the checklist, which was interpreted by the authors as evidence of response generalization. Performance remained stable during follow-up testing. The subsequent three cases showed improvement in self-initiation in vocational tasks as a result of the use of checklists. In each case, the introduction of self-initiation checklists reduced the amount of prompting and increased the number of tasks completed correctly. The results for each patient were obtained with two reversal designs and one multiple baseline design. The third study focused on two patients who exhibited inappropriate sexual behaviour. Improvement of self-regulation was achieved through structured verbal feedback in order to reinforce and shape behaviour and by using a self-monitoring notebook. Basic AB-designs were used to investigate the effect of the interventions. Generalization seems to have occurred, because all six subjects were able to perform their tasks and showed adequate behaviour after withdrawal of treatment.

Ehlhardt, Sohlberg, Glang, and Albin (2005) conducted a pilot study with a multiple baseline across participants design to evaluate an instructional package (TEACH-M) to facilitate the learning and retention of a multi-step task by participants with impaired memory and executive functions. The results showed that four participants were able to learn a 7-step e-mail task following a series of instructions given by a therapist. As a group the participants demonstrated maintenance of effects after 30 days. All participants demonstrated generalization of treatment effects to an altered
administered twice (Manly et al., 2004). One of the single-case studies involved 15 participants (Wilson et al., 1997), so that the effect for the group as a whole could be investigated. This improves internal validity in terms of replication across subjects. Treatment phases were replicated in three single-case studies (Evans et al., 1998; Fish, 2008b; Kirsch et al., 1992). Burke et al. (1991) used multiple baseline designs in two cases. With this design it is possible to rule out alternative explanations for changes in behaviour. The multiple baseline design used by Ehlhardt et al. (2005) also enabled replication of treatment effects across four subjects. In 15 studies the trained tasks were naturalistic tasks in that they were similar to or true real-life tasks, or tasks that were practiced in a natural environment (e.g., home or work), conditions under which generalization is more likely to be achieved. However, in 18 studies training results were obtained directly from the trained task (i.e., the trained tasks and outcome measures were the same). In order to ascertain that the intervention was effective, and not merely a practice effect, various other, additional untrained tasks have to be used as outcome measures. On the other hand, the aim of several of these intervention studies was to improve a particular skill and generalization was not a primary objective (e.g., to investigate the merits of a pager, checklist, prompts, or explicit instructions, without further intention). In one study the outcome measures consisted (also) of neuropsychological tests. However, generalization of treatment effects to daily life cannot easily be inferred from neuropsychological tests. Generalization was specifically addressed in eight studies (although mostly through anecdotal information). Long-term effects of the interventions were assessed in seven studies. All the studies in this category showed that patients with executive impairments are sensitive to the effects of cueing, prompting, direct instruction, and the use of external aids, and that performance in both laboratory and daily tasks can be improved considerably, provided that external structure is offered.

In a replication of the Fasotti et al. (1992) study, Delazer, Bodner and Benke (1998) aimed to improve arithmetical text problem-solving in three chronic TBI patients, using single-case designs and comparing patients’ initial performances with those of five healthy control subjects who received no treatment. Problem-solving performance improved (i.e., increased number of steps to solve the problems and reduction of encoding errors) with systematic cueing. This improvement remained stable after 10 weeks. The training effect did not extend to other tasks than arithmetical text problems (i.e., neuropsychological tests).

Turkstra and Flora (2002) targeted the planning and organizational skills of a TBI patient who was unable to obtain a job because of his poor memory and report-writing skills. He experienced anxiety when faced with the task of taking notes and writing reports. Training consisted of taking notes using a structured format with carrier phrases, while role playing an interview. Generalization was promoted by training tasks that would occur routinely in the workplace. Report-writing accuracy improved significantly. However, no significant differences were found in spelling accuracy or discourse cohesion. Self-reported stress during report-writing was reduced, although stress levels were not measured but only described anecdotally. After training the patient was able to obtain and maintain competitive employment in his chosen profession. It should also be noted that the patient had a history of substance abuse, behavioural problems and multiple traumatic brain injuries.

Conclusion.
In this category three RCTs were found (Man et al., 2006; Soong et al., 2005; Wilson et al., 2001). In the study of Wilson et al. (2001), long-term effects as well as generalization effects were investigated. Also, these effects were demonstrated in three separate subgroups (TBI, stroke and encephalitis). The three quasi-experimental group studies employed non-equivalent control-group designs, in which groups were matched on relevant characteristics, such as IQ, gender, and age, in order to improve internal validity. In the pre-experimental studies care was taken to control for order effects; in one study this was done by randomisation of treatment conditions (Fish et al., 2007), in one by balancing conditions across subjects (Manly et al., 2002) and in one by using a fixed sequence in which with each condition was administered twice (Manly et al., 2004). One of the single-case studies involved 15 participants (Wilson et al., 1997), so that the effect for the group as a whole could be investigated. This improves internal validity in terms of replication across subjects. Treatment phases were replicated in three single-case studies (Evans et al., 1998; Fish, 2008b; Kirsch et al., 1992). Burke et al. (1991) used multiple baseline designs in two cases. With this design it is possible to rule out alternative explanations for changes in behaviour. The multiple baseline design used by Ehlhardt et al. (2005) also enabled replication of treatment effects across four subjects. In 15 studies the trained tasks were naturalistic tasks in that they were similar to or true real-life tasks, or tasks that were practiced in a natural environment (e.g., home or work), conditions under which generalization is more likely to be achieved. However, in 18 studies training results were obtained directly from the trained task (i.e., the trained tasks and outcome measures were the same). In order to ascertain that the intervention was effective, and not merely a practice effect, various other, additional untrained tasks have to be used as outcome measures. On the other hand, the aim of several of these intervention studies was to improve a particular skill and generalization was not a primary objective (e.g., to investigate the merits of a pager, checklist, prompts, or explicit instructions, without further intention). In one study the outcome measures consisted (also) of neuropsychological tests. However, generalization of treatment effects to daily life cannot easily be inferred from neuropsychological tests. Generalization was specifically addressed in eight studies (although mostly through anecdotal information). Long-term effects of the interventions were assessed in seven studies. All the studies in this category showed that patients with executive impairments are sensitive to the effects of cueing, prompting, direct instruction, and the use of external aids, and that performance in both laboratory and daily tasks can be improved considerably, provided that external structure is offered.
<table>
<thead>
<tr>
<th>Study design</th>
<th>Participants</th>
<th>Time since injury</th>
<th>Control group</th>
<th>Intervention</th>
<th>Trained tasks</th>
<th>Outcome measures</th>
<th>Results</th>
<th>Follow-up assessment</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson et al. (2001)</td>
<td>Randomized controlled crossover design</td>
<td>Various aetiologies (143)</td>
<td>M=4.9 years Both groups received same treatment</td>
<td>NeuroPage: portable pager that sends reminders to its users</td>
<td>Everyday activities chosen by participants</td>
<td>Improvement of target behaviours during 3 time periods in which they were monitored</td>
<td>121 / 143 participants were significantly more successful in achieving targets with pager than without</td>
<td>7 weeks after returning pager</td>
<td>Gains maintained stable in 54 participants after returning pager</td>
</tr>
<tr>
<td>Soong et al. (2005)</td>
<td>Randomized controlled design</td>
<td>Unspecified brain injury (15)</td>
<td>No report Random allocation to one of three delivery modes</td>
<td>Interactive skill-training program on solving daily life problems using analogies delivered in one of three modes (online, computer-assisted, or therapist-administered)</td>
<td>Problem solving analogies of daily life problems</td>
<td>HRTB, Lawton IADL, Self-efficacy questionnaire</td>
<td>Improvement on HRTB, IADL, and self-efficacy for all three groups</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Man et al. (2006)</td>
<td>Randomized controlled design</td>
<td>Various aetiologies (83)</td>
<td>Less than six months No treatment group: various aetiologies (20)</td>
<td>Interactive skill-training program on solving daily life problems using analogies delivered in one of three modes (online, computer-assisted, or therapist-administered)</td>
<td>Problem solving analogies of daily life problems</td>
<td>HRTB, Lawton IADL, Self-efficacy questionnaire</td>
<td>Significant within-group improvement in 3 training groups in problem-solving abilities; no significant differences between trained groups on HRTB, IADL, or self-efficacy. No significant differences between treatment groups and control group.</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Fasotti et al. (1992)</td>
<td>Controlled group study</td>
<td>Frontal lobe damage (12 TBI, 13 vascular, 15 tumour)</td>
<td>No report Healthy controls (10) posterior hemisphere lesions (10): cueing procedure</td>
<td>Cuing procedure: specific questions to improve text encoding skills to ameliorate arithmetical-word problem solving</td>
<td>Cued and uncued arithmetical word problems</td>
<td>Number of operations per problem and per series; time needed to solve problems</td>
<td>Improved performance in bilateral frontal and left frontal patients; effect of cueing on solution time only in bilateral frontal group</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Manly et al. (2002)</td>
<td>Controlled group study</td>
<td>TBI (9), CVA (1)</td>
<td>Range 13-124 months (M= 54.4) Healthy controls (24): no-cueing condition</td>
<td>Brief auditory cueing during task performance: facilitate link between intention and action</td>
<td>Cued and uncued performance of Hotel Task (six elements planning task)</td>
<td>Hotel Task</td>
<td>Increased number of tasks attempted, improved time allocation efficiency; no improvement on prospective memory or in time monitoring</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Manly et al. (2004)</td>
<td>Controlled group study</td>
<td>CVA (7)</td>
<td>M= 41.0 months Healthy controls (7)</td>
<td>Brief auditory cueing during task performance: improve maintenance of attentive control</td>
<td>Cued and uncued performance of SART</td>
<td>SART: errors of commission, errors of omission, reaction times</td>
<td>Patients performed worse on SART than controls; performance improved with cueing in both groups;</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Study design</td>
<td>Participants</td>
<td>Time since injury</td>
<td>Control group</td>
<td>Intervention</td>
<td>Trained tasks</td>
<td>Outcome measures</td>
<td>Results</td>
<td>Follow-up assessment</td>
<td>Generalization</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>---------</td>
<td>---------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Manly et al. (2004) -2</td>
<td>Within subjects design</td>
<td>Healthy controls (30)</td>
<td>2</td>
<td>Brief auditory cueing during task performance: improve maintenance of attentive control</td>
<td>Uncued, contingent cued, and random cued performance of SART</td>
<td>SART: errors of commission, errors of omission, reaction times</td>
<td>Improved performance in condition with contingent cues and random cues</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Manly et al. (2004) -3</td>
<td>Within subjects design</td>
<td>Healthy controls (30)</td>
<td>3</td>
<td>Brief auditory cueing during task performance: improve maintenance of attentive control</td>
<td>Cued random sequence SART and cue fixed-sequence SART</td>
<td>SART: reaction times, errors of commission</td>
<td>Effects of cueing were greater in fixed sequence performance</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Fish et al. (2007)</td>
<td>Within subjects design</td>
<td>TBI (14), CVA (4), SAH (1), hypoxia (1)</td>
<td>Median= 27.5 months</td>
<td>Content-free cueing by means of text message sent to mobile telephone, to improve prospective memory performance</td>
<td>Make telephone calls at 4 specified times each day for 15 days, of which 5 days were baseline; 10 days were randomly assigned to either cued or uncued condition</td>
<td>Telephone task proportion of calls and timing accuracy</td>
<td>Significantly improved performance on cued days</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Wilson et al. (1997)</td>
<td>Single-case design (A-B-A)</td>
<td>Various aetiologies (15)</td>
<td>Range 6 months – 13 years</td>
<td>NeuroPage: portable pager that sends reminders to its users</td>
<td>Everyday activities chosen by participants</td>
<td>Achievement of target behaviours at baseline, during pager use and after returning pager</td>
<td>15 subjects benefited; group as a whole was more successful with pager than without</td>
<td>No report</td>
<td>For some subjects there was no decline in performance after pager was withdrawn</td>
</tr>
<tr>
<td>Evans et al. (1998)</td>
<td>Single-case design (A-B-A-B)</td>
<td>SAH (1)</td>
<td>7 years</td>
<td>NeuroPage: portable pager that sends reminders to its users; use of checklist</td>
<td>Everyday activities chosen by participant</td>
<td>Achievement of target behaviours with pager and checklist</td>
<td>Subject improved in achieving daily tasks using pager and checklist</td>
<td>No report</td>
<td>Success restricted to trained tasks; without pager performance back to baseline levels</td>
</tr>
<tr>
<td>Fish et al. (2008)</td>
<td>Single-case design (A-B-A-B-C/A-D)</td>
<td>SAH (1)</td>
<td>17 years</td>
<td>NeuroPage: portable pager that sends reminders to its users; use of checklist</td>
<td>Everyday activities chosen by participant</td>
<td>Achievement of target behaviours with pager and checklist</td>
<td>Subject performed better with pager than with checklist</td>
<td>6 months: subject continues to use pager, but not checklist</td>
<td>No report</td>
</tr>
<tr>
<td>Wade &amp; Troy (2001)</td>
<td>Single-case design (A-B)</td>
<td>CVA, TBI (3)</td>
<td>Range 1 - 15 years</td>
<td>Specified reminder messages on mobile phone</td>
<td>Everyday activities chosen by participants</td>
<td>Self-initiated performance in remembering to carry out target tasks prior and with mobile phone measured through individualized diaries</td>
<td>Use of the mobile phones effective in all five users; use discontinued in one; one inconsistent phone use and variable diary entries</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Study design</td>
<td>Participants</td>
<td>Time since injury</td>
<td>Control group</td>
<td>Intervention</td>
<td>Trained tasks</td>
<td>Outcome measures</td>
<td>Results</td>
<td>Follow-up assessment</td>
<td>Generalization</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>--------</td>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Kirsch et al. (1992)</td>
<td>Single-case design (A-B-A-B-A; A-B-A)</td>
<td>TBI (4)</td>
<td>Range 1-10 years</td>
<td>Interactive Task Guidance (ITG): facilitate performance on functional tasks through the use of cues that are presented on a computer screen</td>
<td>Janitorial task either with computer assistance or with written directions</td>
<td>Task performance; percentage of tasks completed</td>
<td>2 subjects benefited from computerized intervention over written directions; 1 did not benefit; results of 1 equivocal</td>
<td>No report</td>
<td>No report</td>
</tr>
<tr>
<td>Burke et al. (1991) - 1</td>
<td>Single-case design (multiple baseline across behaviours)</td>
<td>TBI (1)</td>
<td>No report</td>
<td>Formulation and use of task-specific checklist</td>
<td>Vocational tasks</td>
<td>Percentage of tasks completed independently</td>
<td>Performance improved to 100%</td>
<td>Gains stable at 3 months</td>
<td>Performance maintained after withdrawal of checklists; 4th task learned without intervention</td>
</tr>
<tr>
<td>Burke et al. (1991) - 2</td>
<td>Single-case design (A-B-A)</td>
<td>TBI (3)</td>
<td>Range 5 – 22 years</td>
<td>Initiation cues and self-initiation checklists</td>
<td>Vocational tasks</td>
<td>Frequency of cues to complete tasks; percentage of task components completed</td>
<td>Maximum performances with cues and checklists; frequency of cues reduced to near zero</td>
<td>Gains stable up to 3 months</td>
<td>Maximum performance without use of checklist or cues</td>
</tr>
<tr>
<td>Burke et al. (1991) - 3</td>
<td>Single-case design (A-B-A)</td>
<td>TBI (2)</td>
<td>No report</td>
<td>Structured feedback, self-monitoring notebook, role-playing</td>
<td>Interaction with opposite sex</td>
<td>Frequency of sexual behaviours</td>
<td>Inappropriate sexual behaviour reduced to near zero for both subjects</td>
<td>Gains maintained at 30 days / 15 months</td>
<td>No report</td>
</tr>
<tr>
<td>Ehlhardt et al. (2005)</td>
<td>Single-case design (multiple baseline across subjects)</td>
<td>TBI (4)</td>
<td>Range 16 – 31 years (M=23 years)</td>
<td>TEACH-M: Instructional sequence consisting of task analysis, step-by-step errorless learning, assessment of performance, review of learned skills, high rates of practice and spaced retrieval, predicting and checking performance</td>
<td>7-step e-mail task (with simplified interface)</td>
<td>Number of correct e-mail steps completed</td>
<td>All participants learned the e-mail task (maximum increase in number of steps recalled)</td>
<td>Gains maintained at 30 days</td>
<td>Generalization of effects shown with slightly altered e-mail interface. No generalization to memory game. No report of generalization to commercial e-mail system or other multi-step tasks.</td>
</tr>
<tr>
<td>Delazer et al. (1998)</td>
<td>Single-case design (A-B-A)</td>
<td>TBI (5)</td>
<td>Range 1.5 to 2 years</td>
<td>Cueing procedure: specific questions to improve text encoding skills to ameliorate arithmetical-word problem solving</td>
<td>Cued arithmetical-word problems, comparable in complexity to test problems</td>
<td>Neuropsychological tests of executive function; arithmetical-word problems (number of operational steps and errors)</td>
<td>Number of steps to solve problem increased, reduction in number of encoding errors, but not in execution errors; no changes on neuropsychological tests</td>
<td>Improvement stable at 10 weeks</td>
<td>Intervention selective for arithmetical word problem solving, not affecting other 'frontal' functions (measured with neuropsychological tests)</td>
</tr>
</tbody>
</table>
Quasi-experimental group studies. Stablum, Umilta, Mogentale, Carlan, and Guerrini (2000) investigated the ability of patients to co-ordinate two simultaneous actions. The prospect of ameliorating the decision-making process was examined using a computerized dual-task in two experiments, involving 10 patients with closed head injury (CHI), matched with 10 healthy control subjects, and 9 patients with anterior communicating artery aneurysm (ACoA), matched with 9 healthy control subjects in the second experiment. In the experiment with CHI patients, both patients and controls received the dual-task training, whereas in the ACoA-experiment the control group received no treatment. Both the patient groups showed a significant decrease in reaction times on the dual task after training, which remained stable at follow-up after 3 months, and also after 12 months for the ACoA group. However, the dual task was used as both pre- and post-measure and as a training task, so the effects of the dual-task training cannot adequately be assessed. Improvements on neuropsychological tests were also found in both patient groups. In the ACoA-group a reduction of complaints was evidenced using a questionnaire (see table 2) of daily activities. The latter finding is an indication of generalization of training effects. However, the relevance and effect of the training for daily life was not investigated with (simulations of) daily life tasks.

Restorative Approach

The objective of this approach is to restore or ‘repair’ underlying executive processes. It is assumed that executive processes (e.g., flexibility, working memory, multi-tasking) can be selectively damaged and that these processes can be directly retrained or remediated. Interventions usually consist of repeated computerized exercises or the repeated administration of (variations of) standardized tests. However, therapy exercises may resemble skills training activities, although the expectation is that in addition to training a specific skill (e.g., remembering to check one’s diary), there is a correspondent improvement in the underlying cognitive process (e.g. increased prospective memory span).

Six intervention studies, described in 5 papers, were found in this category. One study was a true experimental study, three were quasi-experimental studies, one was pre-experimental, and one study used a single-case design.

True experimental group studies. Westerberg et al. (2007) conducted a study with 18 chronic stroke patients who were randomly assigned to either an experimental group or a no-treatment control group, to investigate the effect of a computerized training method for working memory impairments. Comparison of pre- and post-training assessments, using a neuropsychological test battery, resulted in a significant improvement for the treatment group on working memory tests. However, it should be noted that at the post-treatment assessment, the test administrators were no longer blind to the study. There was a significant reduction of cognitive symptoms as measured by a self-rating scale in the treatment group, from which it was concluded that training effects generalized to cognitive functioning in daily life. However, conclusions from the study are limited by the low number of participants and the lack of a follow-up assessment. Moreover, the control group was a no-treatment group which tempers the internal validity of the study.

Table 1b

<table>
<thead>
<tr>
<th>Study design</th>
<th>Participants</th>
<th>Time since injury</th>
<th>Control group</th>
<th>Intervention</th>
<th>Trained tasks</th>
<th>Outcome measures</th>
<th>Results</th>
<th>Follow-up assessment</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkstra &amp; Flora (2002)</td>
<td>Case report</td>
<td>TBI (1)</td>
<td>23 years</td>
<td>Use of structured format for written output to achieve independence in vocational report-writing</td>
<td>Role-playing of vocational tasks</td>
<td>Number of facts accurately reported; spelling accuracy; discourse cohesion</td>
<td>Improvement in report accuracy, no change in spelling and discourse cohesion</td>
<td>No report</td>
<td>Patient reported less anxiety and found and maintained competitive employment</td>
</tr>
</tbody>
</table>
| Note. TBI, traumatic brain injury; CVA, cerebrovascular accident; CHI, closed head injury; SAH, subarachnoid haemorrhage; HRB, Halstead-Reitan Test Battery; SART, Sustained Attention to Response Test.
Stablum, Umilta, Mazzoldi, Pastore, and Magon (2007) purported to improve endogenous (i.e., prepared) task shifting in CHI-patients. Twenty severe CHI patients were divided in a treatment group and a control group that underwent assessment and retest only. Eight mild CHI patients also underwent treatment, as well a group of 18 non-brain damaged subjects. The treatment consisted of five sessions of a computerized shifting task. The study confirmed that the CHI patients were slower than healthy controls in co-ordinating the sequence of task execution when confronted with a shift task and that it is possible to reduce shift-time in CHI patients by means of the treatment. It was claimed that generalization took place, because the severe and mild CHI patients who received treatment improved on neuropsychological tests and a computerized dual task. At follow-up after 4 months, the shift cost was significantly greater than at retest, but still significantly smaller than at baseline assessment. Improvement on neuropsychological tests remained stable at follow-up. Improvement in daily life activities was not examined.

Pre-experimental group studies.
Serino et al. (2007) studied 9 TBI patients who underwent Working Memory Training (WMT), consisting of the repeated administration of three variants of the Paced Auditory Serial Addition Test (PASAT). Comparison of performance on neuropsychological tests before and after the training revealed improvement on tests of working memory, divided attention, executive functions, and long-term memory, but not on tests of speed of processing and sustained attention, which was explained as a verification of the selectivity of the WMT. A single-group repeated measures design was used, in which the TBI group was successively treated in the control condition and in the experimental condition. This design entails the possibility of carry-over effects from one treatment condition to the next, so that the improvements on the neuropsychological tests may be spurious. This threat to the internal validity was not controlled through counterbalancing of treatment conditions. Moreover, it remains unclear for which tests parallel versions were used in order to minimize practice effects. Two general questionnaires were administered to assess patients’ everyday functioning (see table 2), in which improvement was found after the first (control) and the second (experimental) treatment condition. This result suggests that attributing generalization to the WMT is questionable. There was no follow-up assessment.

Single-case studies.
Sohlberg, White, Evans, and Mateer (1992) have sought to increase the time interval between encoding and execution of intentions in order to improve prospective memory in a brain-injured patient. The results of their study showed a steady increase in the length of time between task administration and execution. Task performance was better in the A-condition in which time levels were only slightly above the subject’s prospective memory threshold, whereas execution timing was better in the B-condition, in which time levels were much greater than the patient’s current measured prospective memory ability. At the beginning and in the middle of each phase of the study, generalization probes were administered, i.e., daily life prospective memory tasks and standardized recall tests. Performance on these generalization probes improved in the A-condition, but not in the B-condition. Generalization effects to retrospective memory ability did not show a stable improvement over time. Although the training in general resulted in an improvement of prospective memory, the changes were not experimentally controlled. The authors acknowledge this is a result of flaws in the study design. Follow-up measurements were not included in this study.

Conclusion.
In the restorative approach, one true experimental study was identified (Westerberg et al., 2007). In this study, the random allocation of patients to conditions allows for inferences about the effects of the treatment on working memory. Generalization of effects was established by means of a questionnaire, which showed a reduction of cognitive complaints. However, a general questionnaire does only partially measure generalization as defined in this paper. In the quasi-experimental group studies (Stablum et al., 2000, 2007) threats to internal validity were controlled by matching subjects on relevant characteristics. Also, care was taken to expose all groups to similar treatment conditions. Follow-up measurements were given in three studies and showed no decline in performance. Generalization was demonstrated by performance on tasks other than the trained tasks in four of the experiments and with questionnaires in three experiments. However, the tasks used were various neuropsychological tests some of which are known to have limited ecological value (Burgess et al, 2006; Manchester, Priestley, & Jackson, 2004). Only the study on prospective memory involved naturalistic tasks (Sohlberg et al., 1992). In one study generalization of treatment effects was reported as anecdotal information (Stablum et al., 2000). In short, the use of generalization measures in this category of studies was generally limited. As mentioned before, generalization is expected to occur in the restorative approach, and the results of this review show that improvement of restored cognitive processes is not always measured at the level of improvement in daily life situations.
<table>
<thead>
<tr>
<th>Study design</th>
<th>Participants</th>
<th>Time since injury</th>
<th>Control group</th>
<th>Intervention</th>
<th>Trained tasks</th>
<th>Outcome measures</th>
<th>Results</th>
<th>Follow-up assessment</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westerberg (2007)</td>
<td>Randomized controlled design</td>
<td>CVA (18)</td>
<td>Range 12-36 months (M=20.1)</td>
<td>Patient control group (9): no treatment</td>
<td>Computerized working memory exercises including feedback on performance</td>
<td>Tests of working memory, attention, reasoning, inhibition, and declarative memory; self-rating scale of cognitive failures (CFQ)</td>
<td>Improvement on tests of working memory and attention; no improvement on tests of reasoning, inhibition, or declarative memory; reduction of cognitive symptoms on self-rating</td>
<td>No report</td>
<td>Reduction of cognitive symptoms on self-rating (CFQ)</td>
</tr>
<tr>
<td>Stablum et al. (2000) –1</td>
<td>Controlled group study</td>
<td>CHI (10)</td>
<td>Range 7-61 months (M=27.8)</td>
<td>Healthy controls (10): Dual Task (same as experimental group)</td>
<td>Repetitions of Dual Task: reduction of dual-task cost (reaction times)</td>
<td>Dual Task, PASAT</td>
<td>Improvement on Dual Task reaction times and on PASAT</td>
<td>Gains on Dual Task stable at 3 months</td>
<td>Improved PASAT performance; anecdotal report of generalization real-life situations</td>
</tr>
<tr>
<td>Stablum et al. (2000) –2</td>
<td>Controlled group study</td>
<td>ACoA aneurysm (9)</td>
<td>Range 2-7 months (M=3.66)</td>
<td>Healthy controls (9): no treatment</td>
<td>Repetitions of Dual Task: reduction of dual-task cost (reaction times)</td>
<td>Dual Task, PASAT, CPT, CFQ</td>
<td>Improvement on Dual Task reaction times and on PASAT</td>
<td>At 12 months: gains stable on Dual Task, improvement on tests of executive function; improvement on CFQ</td>
<td>Improved PASAT and CPT performance; Improvement on self-assessment of daily activities (CFQ)</td>
</tr>
<tr>
<td>Stablum et al. (2007)</td>
<td>Controlled group study</td>
<td>CHI severe (20), mild (8)</td>
<td>Range 6-264 months (M=3.66)</td>
<td>severe CHI (10): assessment only; non-brain damaged controls (18); Task Shift (same as experimental group)</td>
<td>Repetitions of Task Shift: reduction of shift cost (reaction times)</td>
<td>Computerized Task Shift: endogenous and exogenous task shift</td>
<td>Task Shift, Dual Task, PASAT, BADS</td>
<td>Improvement on Task Shift reaction times, Dual Task, PASAT, BADS</td>
<td>Gains on Task Shift and BADS stable at 4 months</td>
</tr>
<tr>
<td>Serino et al. (2007)</td>
<td>Within subjects design</td>
<td>TBI (9)</td>
<td>Range 6-61 months (M=28)</td>
<td>Working Memory Training (WMT): recovering CES deficits vs. General Stimulation Training (control training)</td>
<td>Repeated administration of variants of the PASAT</td>
<td>Neuropsychological tests (speed processing, sustained and divided attention, working memory, long term memory, executive functions); Questionnaires of psychosocial functioning</td>
<td>Improvement on tests of working memory, divided attention, executive functions, long term memory; Improvement on psychosocial outcome</td>
<td>No report</td>
<td>Improvement on PCRS and RHFUQ (i.e., competency in daily life)</td>
</tr>
</tbody>
</table>
randomly allocated to a treatment group or a waiting list group. A significant decrease in anger was found in the treatment group, as measured by a specific anger inventory. Unfortunately, anger logs that were completed by participants to monitor the frequency and intensity of anger could not be analysed, due to large numbers of missing data values. Gains in the management of anger were maintained at follow-up after 2 months. Significant generalization of treatment effects to anxiety, depression, self-esteem, or degree of self-awareness was not found.

Behaviour modification

Executive dysfunction may encompass behavioural and emotional disturbances, such as a defective capacity for self-inhibition and self-monitoring in social situations. These impairments can be summarized under the heading of problems with social cognition. Several interventions have been developed to manage the behavioural and emotional changes that can be attributed to executive dyscontrol. Behaviour problems, such as aggression, verbal outbursts with shouting and swearing, attention-seeking behaviour, making inappropriate comments, and repetitive speech, can hamper the learning of new skills and may lead to avoidance of rehabilitation activities. Interventions in this category aim to decrease the frequency of undesired and inappropriate responses or increase the frequency of appropriate responses. Interventions are usually based on learning principles and behaviour therapy.

Twelve studies were identified that evaluated the effects of behaviour modification techniques. One study was a true experimental group study. All the other studies employed single-case designs.

**True experimental group studies.**

The only group study in this category investigated an anger management programme (Medd & Tate, 2000). Sixteen TBI and stroke patients were randomly allocated to a treatment group or a waiting list group. A significant decrease in anger was found in the treatment group, as measured by a specific anger inventory. Unfortunately, anger logs that were completed by participants to monitor the frequency and intensity of anger could not be analysed, due to large numbers of missing data values. Gains in the management of anger were maintained at follow-up after 2 months. Significant generalization of treatment effects to anxiety, depression, self-esteem, or degree of self-awareness was not found.

**Single-case studies.**

In a series of single-case studies conducted by Alderman and colleagues (Alderman, 1991; Alderman & Burgess, 1994; Alderman, Fry, & Youngson, 1995; Alderman & Knight, 1997; Alderman & Ward, 1991; Knight, Rutherford, Alderman, & Swan, 2002) several interventions were described that aimed to modify behaviour problems of patients with severe executive impairments within a clinical setting. The method of satiation through negative practice proved to be effective in a patient who had learned that by shouting he could avoid participating in physically strenuous activities, thereby further exacerbating his problems. His behaviour appeared to have been negatively reinforced. The patient was prompted to keep shouting and this brought about a significant reduction in the undesirable behaviour, enabling him to
participate in rehabilitation activities (Alderman, 1991). Response Cost (RC) is a procedure in which something that is of value to the patient is removed every time a defined target behaviour occurs. RC appeared to be effective in reducing the frequency of several behaviours, such as repetitive speech, swearing and shouting (Alderman & Burgess, 1994; Alderman et al., 1995; Alderman & Ward, 1991). RC was more effective than the use of a token economy system and time-out procedures (Alderman & Burgess, 1994). The effect of RC was strengthened when a procedure called ‘cognitive overlearning’ was introduced, compared to RC alone, resulting in a lower rate of the target behaviour (Alderman & Ward, 1991). While RC had been used to considerable advantage in the rehabilitation unit, it was not an appropriate technique to use outside the ward, because interventions based on this technique require staff to monitor the patients’ behaviour constantly.

An alternative method that may overcome practical limitations of RC (e.g., the number of tokens needed to control high-frequency behaviour disorders), but facilitates learning in the same way, is Self Monitoring Training (SMT). This technique aims to facilitate the patient to self-monitor the frequency of a target behaviour. Although SMT relies on positive reinforcement, is less intrusive, and therefore has fewer ethical constraints, its effect was achieved more slowly than RC, which had an immediate effect on inhibiting the target behaviour (Alderman et al., 1995). Another behavioural technique that is less intrusive than interventions based on punishment and that can be used to reduce behaviours that are incompatible with rehabilitation, is Differential Reinforcement of Low rates of responding (DRL). DRL can be used in cases in which inappropriate behaviour is present in very high frequencies. In DRL, low rates of occurrence of this behaviour are reinforced. A specified target is not to be exceeded by the patient in order to earn reinforcement. The target is subsequently decreased to reduce the behaviour further. DRL methods proved to be effective in three individuals with a range of aggressive behaviours driven by the need to avoid or escape rehabilitation activities (Alderman & Knight, 1997). A DRL programme eventually incorporating SMT, showed improvement of selective attention in two concurrent tasks. SMT alone also led to improvements in selective attention. However, the results of a comparison between both methods remain unclear (Knight et al., 2002).

All the above interventions were investigated by using single-case reversal designs, although one study used a multiple baseline across behaviours design (Alderman & Knight, 1997). Multiple baseline designs and reversal designs come closest to being able to demonstrate causal relationships. Follow-up data were available in one study (Alderman & Ward, 1991) and showed that the treatment gains remained stable over time and outside the training sessions. There was no information available on follow-up measures or of generalization of treatment effects outside the training sessions or environment in the study of Alderman and Burgess (1994). Alderman, et al. (1995) presented follow-up data for the RC intervention that showed generalization of the treatment effects outside the ward. Treatment gains were maintained 15 months after training. However, during this follow-up period SMT was introduced to the same patient. It is unclear how much the therapy contributed to the longevity of increased inhibitory control. Follow-up information was available in the study of Alderman and Knight (1997), showing that the increase in inhibitory control was maintained after DRL had been withdrawn. In the study of Knight et al. (2002), follow-up data were presented in two of the cases which demonstrated long-term effects of the intervention.

In a study by Dayus and van den Broek (2000), a patient with delusional confabulations was treated with self-monitoring training for his excessive swearing and references to delusional confabulations, invoking an ABA design. Results show that following implementation of SMT the frequency of swearing decreased, while the accuracy of self-monitoring improved. This remained stable at follow-up after 3 months. Qualitative information revealed that reduction in swearing and delusions had generalized to everyday situations.

A study by Hanlon, Clontz, and Thomas (1993) targeted four disturbing behaviours in one patient by means of DRI (conditioning of incompatible responses), using separate ABAB designs for each behaviour. Although frequencies of target behaviour are visualised in graphs, and seem to decline over time, no statistical analyses are presented. Apart from anecdotal information, no follow-up data are available.

Matthey (1996) used self-monitoring and extinction techniques in order to modify perseverative behaviour in a patient with anoxic brain damage, who made persistent requests to use the telephone. The treatment was given in two stages and involved feedback, and feedback plus extinction. Although the problem behaviour extinguished after the introduction of both stages of treatment, it did not generalize to the home environment following the patients’ discharge from hospital. The authors acknowledge that training of family members to implement the intervention at home was insufficient. The study design and the lack of statistical analysis make it impossible to assert unequivocally that the interventions were causing the changes.

A single-case study by Watson, Rutterford, Shortland, Williamson, and Alderman (2001) demonstrated the effective use of DRL in reducing the
frequency and severity of aggressive behaviour in one patient 10 years after a severe TBI had been sustained. This study can be considered as a replication of an earlier studies on DRL (Alderman & Knight, 1997; Knight et al., 2002). Moreover, the fact that improvement was achieved within an environment that was more typical of a standard neurorehabilitation unit than a specialized neurobehavioural service, and that treatment was continued in a group home, demonstrates that it is possible to generalize the effects of DRL. This study was not followed up.

Manchester, Hodgkinson, and Casey (1997) discussed the need for specialized units dealing with severe behavioural problems following head injury. These authors described the case of a patient who presented with verbal and physical aggression. Aggressive incidents and shouting were reduced and personal care and communication skills were improved by means of implementation of various reinforcement procedures and the withdrawal of medication. At follow-up after 10 months no further aggressive outbursts were noted, systematic reward of behaviour and personal care tasks was no longer required, and the patient was able to perform household tasks.

Conclusion.
The only group study within this category of behaviour modification interventions was conducted by Medd and Tate (2000). Although this was a preliminary study in which 16 patients participated, a randomized controlled design was used. Positive results were found for the treatment group that were maintained over time. The remaining studies in this category were single-case studies. Of the 11 single-case studies described here, four used a basic AB design (Alderman, 1991; Manchester et al., 1997; Matthey, 1996; Watson et al., 2001) and one used a multiple baseline across behaviours design (Alderman & Knight, 1997). The remaining studies employed reversal designs. It can be concluded that overall, efforts were made to control for confounding variables, although this was done to variable degrees. Several interventions were repeatedly investigated: the effects of DRL, DRI, SMT and RC were the subject of study three times. Such replications enhance the level of evidence of these interventions. Follow-up assessment took place in eight of the studies and in all of these studies treatment gains were maintained. The interventions were highly specific and tailored to the patient. The most obvious limitation of these studies is their restricted application beyond the individuals studied. In the group study the degree of generalization to daily life situations was assessed by means of questionnaires. In some of the single-case studies anecdotal information on generalization to other settings was available. In five studies this information was gathered outside the treatment setting, whilst in two studies treatment continued outside the rehabilitation ward (i.e., at home; Table 3).
<table>
<thead>
<tr>
<th>Study design</th>
<th>Participants</th>
<th>Time since injury</th>
<th>Intervention</th>
<th>Trained tasks</th>
<th>Outcome measures</th>
<th>Results</th>
<th>Follow up assessment</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medd, &amp; Tate (2000)</td>
<td>Randomized controlled crossover design. Patient control group: waiting list.</td>
<td>TBI, CVA (16)</td>
<td>M= 37.2 months</td>
<td>Anger management</td>
<td>No report</td>
<td>Anger logs; inventories of anger (STAXI), self-esteem (SEI), self-awareness (PCRS), anxiety and depression (HADS).</td>
<td>Anger logs unable to be analysed; decrease on anger inventory for treatment group</td>
<td>Anger stable after 2 months</td>
</tr>
<tr>
<td>Alderman (1991)</td>
<td>Single case design (A-B-C)</td>
<td>TBI (1)</td>
<td>6 years</td>
<td>Satiation through negative practice vs. medication</td>
<td>Rehabilitation activities</td>
<td>Frequency and duration of shouting</td>
<td>Frequency and duration of shouting decreased significantly with satiation treatment, not with medication</td>
<td>No report</td>
</tr>
<tr>
<td>Alderman &amp; Ward (1991)</td>
<td>Single-case design (A-B-A-C-A-CD)</td>
<td>HSE (1)</td>
<td>1 year</td>
<td>Response cost, Cognitive overlearning</td>
<td>Rehabilitation activities</td>
<td>Frequency of repetitive speech during therapy sessions</td>
<td>Mean rates of repetitions for 3 treatment conditions lower than baseline conditions. Cognitive overlearning resulted in lower rate of repetition compared to response cost.</td>
<td>Treatment gains maintained at 1 week and 3 months</td>
</tr>
<tr>
<td>Alderman &amp; Burgess (1994)</td>
<td>Single-case design (A-B-A-B)</td>
<td>HSE (1)</td>
<td>32 months</td>
<td>Token economy, DRI</td>
<td>Rehabilitation activities</td>
<td>Frequency of swearing, verbal abuse and rhyming during therapy sessions</td>
<td>Token economy and DRI not effective. Response cost effective in reducing target behaviour</td>
<td>No report</td>
</tr>
<tr>
<td>Alderman et al. (1993)</td>
<td>Single-case design (A-B-C-D-E)</td>
<td>HSE (1)</td>
<td>13 months</td>
<td>Response cost, SMT</td>
<td>Rehabilitation activities</td>
<td>Frequency of verbal intrusions, percentage agreement between patient and therapist</td>
<td>Target behaviour reduced with extended intervention, improved accuracy of self-monitoring</td>
<td>Gains stable at 5 months</td>
</tr>
<tr>
<td>Alderman &amp; Knight (1997)</td>
<td>Single-case design (multiple baseline across behaviours; A-B-A-C; A-B)</td>
<td>TBI, SAH (3)</td>
<td>2 years; 3 years; 7 years</td>
<td>DRL</td>
<td>Rehabilitation activities</td>
<td>Frequency of physical aggression, verbal outbursts, sexual comments, and swearing</td>
<td>Reduction in target behaviours in all patients</td>
<td>Case 1: gains stable at 1, 4, 12 weeks and 18 months; Case 2: gains stable at 5 months; Case 3: no report</td>
</tr>
<tr>
<td>Knight et al. (2002)</td>
<td>Single-case design (A-B-A; A-B-C-A; A-B-C-D-E)</td>
<td>TBI, CVA (3)</td>
<td>5 years; 5 years; 13 years</td>
<td>SMT, DRL</td>
<td>Rehabilitation activities</td>
<td>Frequency of provocative and perseverative comments</td>
<td>DRL resulted in most rapid change, SMT facilitates greater improvements in selective attention</td>
<td>Case 1: gains stable after 2 years; Case 2: gains stable at 4 weeks and 5 months; Case 3: gains stable at 1 and 3 months</td>
</tr>
<tr>
<td>Dayus &amp; van den Broek (2000)</td>
<td>Single-case design (A-B-A)</td>
<td>SAH (1)</td>
<td>6 years</td>
<td>SMT</td>
<td>Structured interview (conversation)</td>
<td>Frequency of swearing and references to delusions; agreement score between patient, therapist and spouse</td>
<td>Improved accuracy; references to delusions reduced</td>
<td>Gains maintained at 3 months</td>
</tr>
</tbody>
</table>
were pre-experimental studies without control groups. The majority of studies were single-case studies: 31 out of 54 (57%), 13 of which included more than one patient undergoing the same intervention (case series), thereby providing replication across subjects. Fifteen single-case studies (48%) used reversal designs (with at least one reversal), there were 5 multiple baseline designs (16%), and 1 alternating treatment design. These designs allow a better controlled examination of cause and effect relationships. The remaining single-case studies used a simple AB design or pre-test-post-test designs that do not provide sufficient control for threats to the internal validity. In 37 studies (69%) the trained tasks were ‘naturalistic’ tasks or situations, i.e. (simulations of) daily life tasks. In 45 of the 54 studies (83%) the outcome measures included performance on trained tasks. These latter results show

Discussion

In this review, intervention studies of executive impairments published during the last two decades were evaluated in terms of their clinical effectiveness, by focussing specifically on internal and external validity. A total of 46 papers, describing 54 studies, was included. Of the 54 studies, 12 (22%) were true experimental studies, 6 (11%) were quasi-experimental group studies that used matched control subjects (not randomly allocated), and 5 (9%) were pre-experimental studies without control groups. The majority of studies were single-case studies: 31 out of 54 (57%), 13 of which included more than one patient undergoing the same intervention (case series), thereby providing replication across subjects. Fifteen single-case studies (48%) used reversal designs (with at least one reversal), there were 5 multiple baseline designs (16%), and 1 alternating treatment design. These designs allow a better controlled examination of cause and effect relationships. The remaining single-case studies used a simple AB design or pre-test-post-test designs that do not provide sufficient control for threats to the internal validity. In 37 studies (69%) the trained tasks were ‘naturalistic’ tasks or situations, i.e. (simulations of) daily life tasks. In 45 of the 54 studies (83%) the outcome measures included performance on trained tasks. These latter results show
that subjects have learned a repeatedly practised task, and therefore outcomes mainly reflect task-specific practice-effects. Generalization tasks are needed to evaluate whether performance remains the same in an untrained task. Only 9 (17%) of the intervention studies used such a new, untrained, naturalistic task, specifically designed to measure treatment effects in a new situation or under different circumstances that resemble the demands of daily life. In 22 studies (41%), neuropsychological tests, ‘paper and pencil’ tasks or computerized tasks were used to measure the effects of the intervention. In 14 studies (26%) self-reports, such as questionnaires assessing complaints or well-being, diaries, or Goal Attainment Scaling were used as outcome measures. In 9 studies (17%) observations or ratings by others (e.g., staff members or family) were used as an outcome measure. Long term follow-up assessments were performed in 25 studies (46%). In 34 studies (63%) generalization to other tasks or situations than the ones the patients were trained in, was explicitly mentioned. However, 53% of the studies claiming that generalization took place, reported anecdotal information.

The results of most of the intervention studies seem encouraging, but the methodological rigour of a number of studies can be improved. Only a few studies contain all the essential ingredients necessary to draw valid conclusions about treatment effectiveness, such as control groups, randomization, evidence of real-life generalization and long-term follow-up. Hence, the extent to which behavioural changes can be attributed to the treatment and the degree to which patients benefited from the treatment in their daily lives is not always unequivocal.

Internal validity can be improved if the chosen design sufficiently controls for alternative explanations that may influence performance on outcome measures. Double blind RCTs contain the strongest experimental control and therefore represent the ‘gold standard’ in clinical outcome research (Nezu & Maguth-Nezu, 2008). The results of the present review however, showed that so far RCTs in the field of executive dysfunction rehabilitation are scarce. Apparently, the prospects for conducting a full RCT are limited. These trials are laborious, time-consuming and expensive and therefore not always feasible in clinical practice. The RCTs that we identified showed positive findings. The majority of studies evaluated were single-case studies, which are more easily designed and conducted. The internal validity in single-case designs can be improved in several ways. By replicating experimental phases (e.g., ABAB design) the effect of an experimental treatment can be underpinned more strongly if the pattern of responses reverts back to the original pre-treatment level when the pre-treatment conditions are reinstated (i.e., return to baseline phase; e.g., Knight et al., 2002). If reversal to baseline conditions is not likely to occur, another design, such as a multiple baseline design, may be used to prove the effectiveness of the treatment (e.g., Alderman & Knight, 1997; Burke et al., 1991). Also, the evidential value of a treatment will be stronger if the effect of an intervention can be replicated in more than one patient (e.g., Wilson et al., 1997). The necessary design requirements for strong single-case experimental designs can be found in Tate et al. (2008).

To investigate whether an intervention is effective, independent replications of the treatment’s efficacy are required. The results of this review show that there are only few replication studies, yet in three out of the four treatment approaches that we distinguished in this paper, interventions could be identified that provided stronger evidence in terms of replication. With regard to internal strategies replications were available for Goal Management Training (Levine et al., 2000; Schweizer et al., 2008; Spikman et al., 2010) as well as Problem Solving Training (Von Cramon & Mathes-von Cramon, 1994; Von Cramon et al., 1991; Spikman et al., 2010). In the approach using external strategies, there is strong evidence for the effectiveness of NeuroPage, which was investigated in four studies (Evans et al., 1998; Fish, Manly, Emslie et al., 2008a; Fish, Manly, & Wilson, 2008b; Wilson et al., 1997, 2001). In the behavioural interventions DRL (Alderman & Knight, 1997; Knight et al., 2002; Watson et al., 2001), DRI (Alderman & Burgess, 1994; Hanlon et al., 1993; Watson et al., 2001), SMT (Alderman et al., 1995; Dayus & van den Broek, 2000; Knight et al., 2002) and RC (Alderman & Burgess, 1994; Alderman & Ward, 1991; Alderman et al., 1995) were each investigated three times. In the restorative approach, there were no replication studies.

The number of replication studies seems to be limited because the dysexecutive syndrome consists of a variety symptoms (e.g., Burgess & Simons, 2005; Darling, Della Sala, Gray, & Trivelli, 1998), which has led to an equal amount of treatment procedures, each dealing with a different aspect of the dysexecutive syndrome. This hinders the comparability and replication of studies. To reduce this heterogeneity two options are considered. On the one hand, it is possible to distinguish a limited set of clinically relevant executive symptoms on which to concentrate intervention studies. This review highlights that planning in complex (problem-solving) tasks, maintenance of attentive control and prospective memory, dealing with information competition during multi-tasking, and behavioural inhibition in social situations are such core symptoms. Alternatively, evidence from cognitive neuroscience provides the clinician with emerging theoretical constructs of executive functioning that might guide the development of new treatments. In fact, theories inform interventions of what might be
treated and how they might work. Broadly speaking, two kinds of theories can be distinguished, that have different implications for treatment. On the one hand, single account and single construct-led theories imply that damage to a process (e.g., context information) or a construct (e.g., working memory) is responsible for a number of dysexecutive symptoms. These accounts hold the promise that quite different symptoms might share a common cause, and thus treatment might lead to improvement in a wide range of situations (see Burgess & Simons, 2005). An example is Duncan’s theory of disorganization of behaviour associated with ‘goal neglect’ (Duncan et al., 1996). This theory has inspired Robertson (1996) to develop GMT, a rehabilitation protocol for patients with executive problems hampering goal-directed behaviour in general (see Levine et al., 2000). On the other hand, multiple process theories suggest that the executive system is fractionated and made up of several component processes that can be differentially impaired, leading to different patterns of symptoms in patients. Stuss (2009) for example, suggests a theoretical framework for understanding executive impairments in which the first step is to relate deficits to focal frontal lobe damage. A second step in this line of reasoning (Stuss, 2009) is to delineate deficits and their consequences for daily life which may help to focus rehabilitation interventions and to test whether such interventions lead to improvement of executive functioning in daily life. Stuss distinguishes four separate categories of disorder which are related to damage in different frontal regions: deficient energization (initiation and sustaining behaviour), executive deficits (task setting, monitoring), behavioural and emotional self-regulation impairment, and disturbed meta-cognition (theory of mind). In the case of multiple process theories, intervention practice might focus on various impaired processes with multidimensional or multifaceted treatment protocols (for an example, see Spikman et al., 2010).

The second main objective of the present review was the evaluation of external validity. An important issue concerning external validity is the level of functioning at which the treatment is effective. Different levels of functioning are distinguished within the World Health Organization (WHO) ICF framework (WHO, 2001), that is the level of impairment (relating to bodily and mental functions), the execution of daily activities and social participation. Rehabilitation is aimed at enhancing the patient’s independence, and thus strives to improve daily functioning and social integration. Preferably, executive dysfunction interventions should concentrate on improving executive functioning in daily life and thus include generalization of strategies, skills or the use of external aids as key elements of the training. According to Cicerone et al. (2000, 2005), relatively few studies have directly evaluated the generalization of treatment effects to everyday situations and behaviours, although several provide evidence to support the practical utility of cognitive rehabilitation. Our study confirms Cicerone’s view that treatments directly addressing community integration, social participation or return to work as an outcome of rehabilitation are sparse. Generalization can be achieved by practicing tasks that are, or at least resemble, tasks of daily living, such as cooking or keeping an appointment. In designing new interventions for executive dysfunction generalization can also be facilitated by devising and training variants of the same daily life tasks and gradually increase demands in these tasks. In the cooking example, one could practice various recipes. Moreover, to further optimise generalization, skills and activities should preferably be practiced in different environments. In the cooking example, cooking could be practised in different kitchens. Overlearning, i.e., prolonged rehearsal of a skill or task, is recommended to support generalization in time. Several studies considered in this review did not measure generalization at task level, considering that they used the improved scores from the trained tasks as outcome variables. In other cases, only subjective or qualitative information was given, by means of self-reports or observations by others. Still other studies have only used neuropsychological tests to assess the effectiveness of an intervention. However, as rehabilitation aims to ameliorate everyday functioning, these latter measures might not be congruent with the intended effects of the intervention or lack the necessary ecological validity. Preferably, task variants and neuropsychological tests should be complemented with structured observations (e.g., Pollens, McBratnie, & Burton, 1988), with (semi-)structured interviews or questionnaires that are cross-referenced by proxies (e.g., Dysexecutive Questionnaire or DEX; Wilson, Alderman, Burgess, Emslie, Evans, 1996). In order to assess the maintenance of effects of a treatment over time, all intervention studies should include follow-up measurements.

Finally, one should not forget the economic arguments. The results of this review show that patients with executive impairments benefit from treatment. However, both clinical practice and research have shown that these patients are slow learners, and therefore treatments tend to be lengthy and laborious. Obviously, the cost of care can be quite high, given the intensity and duration of some treatments. Clinical researchers face the challenge of balancing the shortest possible treatment on the one hand and effectiveness on the other hand, in order to warrant reimbursement. This should be an incentive to conduct more clinical effectiveness research in the field of executive impairment rehabilitation.
References


outcome research: A practical guide to conducting randomised controlled trials for psychosocial interventions. New York: Oxford University Press.


Effects of a multifaceted treatment programme for executive dysfunction after acquired brain injury on indications of executive functioning in daily life

Published in:
Journal of the International Neuropsychological Society 2010; 16 (1), 118-129
Abstract

A multicenter randomized control trial (RCT) was conducted to evaluate the effects of a treatment for dysexecutive problems after acquired brain injury (ABI) on daily life functioning. Seventy-five ABI patients were randomly allocated to either the experimental treatment, multifaceted strategy training for executive dysfunction, or a control treatment, computerized cognitive function training. Assessment took place before, directly after and six months post-treatment. The primary outcome measure, the Role Resumption List (RRL), and two other follow-up measures, the Treatment Goal Attainment (TGA), and the Executive Secretarial Task (EST) were indications of daily life executive functioning. The experimental group improved significantly more over time than the controls on the RRL and attained significantly higher scores on the TGA and EST. We conclude that our treatment has resulted in significant improvements of executive functioning in daily life, lasting at least 6 months post-treatment. Although control patients’ satisfaction and subjective well-being were at the same level, the experimental group had better abilities to set and accomplish realistic goals, to plan, initiate, and regulate a series of real-life tasks, and to resume previous roles with respect to work, social relations, leisure activities, and mobility.

Introduction

Executive dysfunction (ED) is a frequent and disabling consequence of acquired brain injury (ABI), which in most cases impairs the patients’ abilities to function independently in daily life. Executive functions (EF) are those capacities that make persons effective in the real world, allowing them to adapt to new situations and to develop and pursue their life goals in a constructive and productive way (Burgess & Simons, 2005). In fact, EF is an umbrella term which encompasses a broad range of higher order capacities for planning, initiation, regulation, and verification of complex, goal-directed behaviour (Lezak, 1982). EF are subserved by prefrontally driven brain circuits in which other cerebral and cerebellar areas take part (Lichter & Cummings, 2001; Sbordone, 2000). It is acknowledged that ED not only results from lesions directly affecting the prefrontal cortex, but also from (diffuse) injuries elsewhere in the brain affecting these circuits (Andres, 2003; Eslinger & Grattan, 1993; Stuss & Alexander, 2000; Stuss, 2006). ED has been extensively documented in traumatic brain injury (TBI) patients (Bamdad, Ryan, & Warden, 2003; Bennett, Ong, & Ponsford, 2005; Busch, McBride, Curtiss, & Vanderploeg, 2005; Hart, Whyte, Kim, & Vaccaro, 2005) with evidence for even more severe problems in the case of focal frontal damage (Fontaine, Azouvi, Remy, Bussel, & Samson, 1999; Spikman, Deelman, & van Zomeren, 2000). Ample evidence of ED has also been found in other ABI patients, for example stroke (Leskela et al., 1999; Pohjasvareaa et al., 2002; Sachdev et al., 2004), cerebral tumours (Goldstein, Obrzut, John, Ledakis, & Armstrong, 2004; Tucha, Smely, Preier, & Lange, 2000), and post-anoxic encephalopathy (Armengol, 2000; Simo-Guerrero et al., 2004).

Many ABI patients with ED are referred for rehabilitation. As ED hampers the capacities for changing and adapting behaviour to altered situations, it often constitutes a major obstacle to the acquisition of independent living skills and hence to successful community re-entry (Fasotti & Spikman, 2002). Therefore, effective interventions aimed at improving EF in daily life are sorely needed. However, difficulties with learning and applying training principles are inherent to ED, and designing clinically relevant interventions requires the consideration of several factors that make EF such complex functions.

The first of these factors is the heterogeneous construct of EF, encompassing a range of different subfunctions. Ylvisaker (1998), for example, distinguishes the following EF aspects: self-awareness of strengths and needs, realistic and concrete goal-setting, planning the steps to these goals, self-initiating these plans, self-monitoring and evaluating performance according to plan and goal,
self-inhibiting behaviour not leading to the goals set, flexibility and problem solving when situations cannot be dealt with according to plan, and finally, strategic behaviour, transfer of successful behaviours to other situations. These aspects can be differentially impaired, leading to different patterns of EF symptoms in patients. Ideally, clinically relevant treatments should be multifaceted and aimed at improving a comprehensive but finite range of EF. So far, such treatments are sparse (Cicerone et al., 2000, 2005). The majority of studies have been carried out addressing a limited set of EF aspects, like problem-solving (Foxx, Martella, & Marchand-Martella, 1989; von Cramon & Matthes-von Cramon, 1994), goal management (Levine et al., 2000), or self-regulation (Medd & Tate, 2000). Notable exceptions are the studies of Rath, Simon, Langenbahn, Sherr, and Diller (2003), Cicerone and Giacino (1992), and Suzman, Morris, Morris, and Millan (1997), in which several aspects of EF were addressed. In the study by Rath et al., for example, the effects of group therapy aimed at improving emotional self-regulation as well as reasoning in everyday problem-solving situations were investigated.

Another factor is the targeted level of functioning. Interventions should be ecologically valid and optimize behaviour in the real world (Worthington, 2005). This usually involves teaching compensatory strategies, that is, top-down approaches, that can be flexibly adapted and applied to the various executive problems that patients encounter in daily situations (Fasotti & Spikman, 2002).

A third factor concerns the measurement of interventions targeted at several executive aspects in daily functioning. Conventional neuropsychological tests tapping single executive aspects are not likely to uncover these effects. Also, many EF tests [the WCST (Wisconsin Card Sorting Test), Stroop test or Trail Making test] do not fully assess executive abilities required in daily life (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Burgess, Alderman, Forbes, Costello, Coates, & Dawson, 2006; Mountain & Snow, 1993; Spooner & Pachana, 2006). The BADS’s (Behavioural Assessment of the Dysexecutive Syndrome; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) ecological validity, for example, is still debated (Norris & Tate, 2000; Wood & Liotti, 2006). A more general problem is that to tap EF, tests should be complex and new, which they can be only once. This makes them unsuitable for repeated assessment and thus as the only outcome measures for a treatment for ED. Hence, treatment effects should also be measured in terms of improvement on indicators of daily life functioning, which so far are neglected outcome-measures (Cicerone, 2004).

In this study, a newly developed multifaceted treatment of ED is presented and evaluated. Herein ABI patients are trained to cope with everyday executive problems in all eight EF aspects distinguished by Ylvisaker (1998). Such a comprehensive treatment has not been applied before. The training was given to ABI patients who were expected to resume or had already resumed (part of) their previous daily life activities. To control for nonspecific effects of the experimental treatment, it was compared with a computerized cognitive function training (Marker, 1987). Such training programs may have specific effects, especially on attention deficits (Sturm et al., 2002; Sturm, Willmes, Orgass, & Hartje, 1997), but their effectiveness in improving daily life skills is still debated. Therefore, and given the general character of the control training, no specific effects on daily executive activities were expected.

Several outcome measures assessing daily life executive functioning were included. The primary outcome measure pertained to the resumption of social roles. Our hypothesis was that the experimental treatment would significantly improve executive functioning in daily life activities and increase social participation immediately after training. However, because patients were taught to apply compensatory strategies autonomously, we considered the presence of treatment effects at follow-up as even more important.

Method

Study Design and Procedure

The study was set up as a prospective multicenter randomized control trial (RCT) with two patient groups receiving treatment in seven Dutch rehabilitation centers and two academic settings. Previously to the trial, several experienced neuropsychologists received extensive training in the use and application of the experimental treatment protocol. During treatment, these therapists were monitored and given feedback in central meetings taking place every 3 months. The same therapists were responsible for the administration of the control treatment.

Data were obtained according to the ethical regulations of the participating institutions, in compliance with the Helsinki Declaration. Participants eligible for the study had to suffer from Acquired Brain Injury (ABI) of non-progressive nature (i.e., TBI, Stroke or cerebral tumours), with a minimal time post-onset of 3 months. Age had to be between 17 and 70 and participants had to live at home. Candidates had to be referred for outpatient rehabilitation with post-injury dysexecutive problems either reported by themselves or observed by proxies. The signalled problems regarding planning, initiation,
and regulation of complex daily life tasks had to hamper the resumption of previous activities and roles. Patients who gave their informed consent underwent a neuropsychological examination.

Dysexecutive problems were measured by means of the Dysexecutive Questionnaire (DEX; Burgess, Wilson, Evans, & Emslie, 1996). Final inclusion was based on the following criteria: a BADS standard age score in the category ‘low average’ or lower, or a discrepancy between BADS standard age score and IQ (Dutch GIT-short version; Luteijn & van der Ploeg, 1983) of 15 points (1 SD), or standard scores of 2 or lower on the BADS’s most complex subtasks, Six Elements Test and Zoo Map Test.

Exclusion criteria were: severe cognitive comorbidity (i.e., aphasia, neglect, amnestic syndrome, indicated by deficient scores on relevant neuropsychological tests) interfering with treatment, severe psychiatric problems, neurodegenerative disorders and substance abuse.

Suitable candidates were blindly and randomly assigned to either the experimental or the control condition per centre. Balanced assignment (per four patients) took place by lot (two ‘control’ and two ‘experimental’). Lots were drawn blindly by an employee not involved in the study. Excluded patients were offered standard rehabilitation.

In each treatment condition, patients underwent 20-24 one-hour treatment sessions, twice a week, during a 3-month period. At baseline (T0), immediately after treatment (T1) and 6 months post-treatment (T2), an extensive battery of tests and questionnaires was administered by independent assessors who were blind for treatment condition, except for the DEX-therapist version and the Executive Observation Scale (EOS), which were therapist-rated. During the interval between T1 and T2, patients underwent no other treatments. At follow-up, these therapists were allowed to talk with and observe the patients in order to complete DEX and EOS forms. Executive Secretarial Task (EST), BADS, and DEX data were also collected in a group of healthy controls, recruited by means of an advertisement in a local paper. DEX-therapist forms were filled in by the assessor.

Patients
Seventy-five patients were included, underwent the treatment (experimental group: 38; control group: 37) and post-training assessment. Five rehabilitation centres supplied the majority of the patients (24, 15, 15, 10, and 9, respectively). In two other centres after treatment of one patient, the therapist had to withdraw for reasons not related to the study, and could not be replaced.

Table 1 shows the characteristics of both patient groups. These were well-matched, as no differences were found after statistical testing (Mann-Whitney U and χ² tests). At follow-up, three patients in the experimental group did not show up and one control patient dropped out due to logistical problems. Figure 1 shows a CONSORT-diagram in which the flow of participants and attrition after initial enrollment is displayed.

Figure 1 The CONSORT-diagram

Experimental Treatment
The Multifaceted Treatment of Executive Dysfunction’s main objective is improvement of the eight EF aspects of Ylvisaker’s conceptual framework: self-awareness, goal-setting, planning, self-initiation, self-monitoring, self-inhibition, flexibility and strategic behaviour.

Improvement was fostered by teaching patients a comprehensive cognitive strategy, which allowed them to tackle daily life situations in a systematic and structured way. In several stages, this strategy forced patients
home situation was accomplished by using relevant exercises and home assignments. A diary was used as planning and memory aid. The comprehensive multifaceted strategy relied heavily on Goal Management Training (GMT; Levine et al., 2000) and Problem Solving Training (PST; von Cramon & Matthes-von Cramon, 1994). Its underlying idea was universal sub-goaling, derived from cognitive architectures like to formulate intentions and actions explicitly in terms of goals and subgoals (planning) and effectively execute these plans, while monitoring their behaviour. The treatment, described in a standard protocol, was given by an experienced rehabilitation- or neuropsychologist. It could be individually tailored to patients’ specific problems, needs and goals, by varying content and number of sessions (up to a maximum of 24). Transfer of learning to the

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Control group (n = 37)</th>
<th>Experimental group (n = 38)</th>
<th>Sign.</th>
<th>Healthy controls (n = 57)</th>
<th>All patients (n = 75)</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M/median (SD))</td>
<td>43.7/ 44 (14.9)</td>
<td>41.4/ 47 (12.1)</td>
<td>n.s.</td>
<td>47.8 (11.4)</td>
<td>42.5 (13.6)</td>
<td>*</td>
</tr>
<tr>
<td>Education (M/median (SD, range))</td>
<td>4.8/ 5 (1.2, 2.7)</td>
<td>5.2/ 5 (1.0, 3.7)</td>
<td>n.s.</td>
<td>5.2/5 (1.0, 3.7)</td>
<td>5.0/5 (1.1, 2.7)</td>
<td>n.s.</td>
</tr>
<tr>
<td>M/F (%)</td>
<td>65/35</td>
<td>68/32</td>
<td>n.s.</td>
<td>41/59</td>
<td>66/33</td>
<td>**</td>
</tr>
<tr>
<td>Chronicity (months)</td>
<td>47/9/19</td>
<td>71/30</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M/median (SD, range))</td>
<td>(64.1, 4-288)</td>
<td>(105.4, 3-468)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aetiology (%) (TBI/stroke/other)</td>
<td>32.5/54/13.5</td>
<td>55/32/13</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BADS Age score</td>
<td>85.9 (14.3)</td>
<td>90.6 (11.4)</td>
<td>n.s.</td>
<td>102.3 (12.3)</td>
<td>88.3 (13.0)</td>
<td>***</td>
</tr>
<tr>
<td>DEX patient</td>
<td>31.0 (13.7)</td>
<td>32.2 (13.1)</td>
<td>n.s.</td>
<td>18.3 (6.6)</td>
<td>31.6 (13.3)</td>
<td>***</td>
</tr>
<tr>
<td>DEX proxy</td>
<td>32.1 (15.5)</td>
<td>32.1 (14.3)</td>
<td>n.s.</td>
<td>18.1 (9.9)</td>
<td>32.0 (14.8)</td>
<td>***</td>
</tr>
<tr>
<td>DEX therapist</td>
<td>35.7 (11.7)</td>
<td>34.9 (13.3)</td>
<td>n.s.</td>
<td>10.1 (6.5)</td>
<td>35.3 (12.1)</td>
<td>***</td>
</tr>
<tr>
<td>Shortened IQ</td>
<td>109 (14.4)</td>
<td>116.1 (16.8)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory IR</td>
<td>38.6 (10.1)</td>
<td>39.0 (11.5)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory DR</td>
<td>7.9 (3.6)</td>
<td>7.6 (3.3)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop 3/2</td>
<td>1.7 (0.3)</td>
<td>1.6 (0.3)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMT-B/A</td>
<td>2.3 (1.1)</td>
<td>2.1 (1.0)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOL nr correct</td>
<td>10.8 (1.3)</td>
<td>10.6 (1.5)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QOLIBRI satisfac</td>
<td>124.7 (28.9)</td>
<td>120.4 (27.7)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QOLIBRI burden</td>
<td>48.1 (13.1)</td>
<td>54.9 (11.5)</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOS</td>
<td>20.6 (3.9)</td>
<td>20.7 (3.2)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRL</td>
<td>7.8 (3.6)</td>
<td>7.2 (3.0)</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Questionnaires and rating scales at T0 of the two patient groups separately. For the total patient group and healthy controls demographic data, BADS age score, and DEX scores are included. Educational level was indicated on a 7-point scale, with 1 = < 6 years primary school and 7 = university education. Differences between groups were tested with two-tailed results of *t* tests or Mann-Whitney *U* tests. M/F = male/female; TBI = traumatic brain injury; BADS = Behavioural Assessment of the Dysexecutive Syndrome; DEX = Dysexecutive Questionnaire; IR = immediate recall; DR = delayed recall; TMT = Trail Making Test; TOL = Tower of London; EOS = Executive Observation Scale; RRL = Role Resumption List; n.s. = not significant.

*p < .05; **p < .01; ***p < .001
Stage 1, Information and Awareness, addressed Ylvisaker’s EF aspect of self-awareness. This preparatory stage consisted of four to six psycho-educative sessions aimed at improving awareness of executive deficits and enhancing motivation for treatment. Patients were extensively informed about executive problems and their negative consequences for daily life in general and their own lives in particular. This information, together with the results of the neuropsychological assessment, provided the basis for a strengths and weaknesses analysis, which formed the starting point for the treatment. Patients were continually stimulated to monitor and evaluate executive performance during training. Also, they were asked to systematically predict their performance in home assignments. Every next session, these predictions and their fulfilment, together with factors that did or did not help were extensively evaluated. These ‘awareness-exercises’ were continued after stage 1 and incorporated in every subsequent training session.

Stage 2, Goal Setting and Planning, consisted of seven to nine sessions and was aimed at training Ylvisakers’ EF aspects of goal setting and planning in a systematic way. Patients were taught to apply the General Planning Approach (GPA) emphasizing the formulation of intended activities and tasks in terms of goals and steps leading to these goals. Daily life goals had to be verbalized explicitly and concretely in terms of when, where, with whom, what and how long, on a fixed GPA worksheet. Patients learned to formulate concrete steps leading to previously set goals and put these steps in the right order. This was first practiced ‘in vitro’, using scripts of Sirigu (Sirigu, Zalla, Pillon, Grafman, Agid, & Dubois, 1996), and subsequently with goals brought forward by the patients. GPA was systematically shaped by asking critical questions (e.g., do you think it is practical to put step x before step y?, don’t you think that this step can be further split-up?, are you sure that you have allocated enough time to step x?), about goals and steps devised by the patient, urging him to reflect on his plans. When goal setting and planning were adequately mastered, patients were encouraged to anticipate on eventual problems by stimulating them to ask “what if…” questions (e.g., what if you need more time as a consequence of your slowness, what if an ingredient in your meal is not readily available?) and to devise alternative steps or plans, following the same shaping procedure. Finally, without help of the therapist patients had to formulate three concrete goals they wanted to achieve by means of the treatment. These three goals had to originate from patients’ difficulties in daily life executive functioning, without other restrictions. Examples of goals suggested by patients were improving the organization of household chores, planning activities in advance, learning to use public transport facilities to increase mobility, organizing activities with family or friends to improve contacts, improving output and quality of activity in volunteer work, etc.

Not until the planning skills of stage 2 were mastered, was the last stage, Initiation, Execution and Regulation, commenced, tapping Ylvisakers’ EF aspects of self-initiation, self-monitoring, self-inhibition, flexibility/problem-solving, and strategic behaviour. In 9-13 sessions effective execution of plans in “vivo” was addressed. Initiative was facilitated by linking plans to an external device, such as a diary, alarm, PDA (personal digital assistant) or mobile phone, or to a routine activity, such as lunch time or the morning hygiene routine, to prompt the first step. The next steps of execution and monitoring were taught according to GMT. Starting from session 17, PST was introduced in order to address problems that might arise during execution of plans. Examples of such problems were sudden situational changes, ‘open-ended’ situations with multiple choices or overwhelming internal or external conditions (e.g., unexpected time pressure). PST was thus used as a method to cope with “what if…- questions” (see Stage 2) in reality. Patients used fixed worksheets to monitor, report and discuss their performance in the therapy sessions. In order to stimulate generalization the patients were repeatedly told they could use the proposed strategy to cope with all multi-step tasks in daily life and continually encouraged to do so.

A (Dutch) treatment manual can be freely obtained from the first author.

Control Treatment

Control treatment was Cogpack (Marker, 1987), an individually administered computerized cognitive training package consisting of several repetitive exercises. It is aimed at improving general cognitive functioning (like reaction speed, attentional functioning, memory and planning). The program is self-supporting; most tasks can be performed without assistance but a therapist was present to provide support when needed. Task performance was followed by direct feedback from the computer program so that patients could gain insight into their strengths and weaknesses. Improvements over time could be monitored. However, no clues about strategic approaches to the proposed tasks were offered by the therapist. Just like in the experimental treatment, after ten sessions patients were asked to formulate three goals...
they wanted to achieve by means of the training. During the remaining sessions, patients could freely select the exercises that they deemed useful for reaching these goals.

**Measures**

An extensive battery of tests and questionnaires was administered to the patients (see also Boelen, Spikman, Rietveld, & Fasotti, 2009). For this study, the following instruments were relevant.

**Primary outcome measure**

In order to measure executive functioning at a social participation level, The Role Resumption List (RRL; Spikman, Brand, & Brouwer, 2003) was administered three times. Based on a structured interview with the patient, the RRL assesses changes in amount and quality of activities compared to premorbid levels in four daily life domains (vocational functioning; van Zomeren & van den Burg, 1985), social interaction with proxies, leisure activities, and mobility). These domains were rated by an assessor on a 5-point scale (0 = no change, 4 = severe loss of independence), with a total score ranging from 0 to 16. Based on the transcript of the interview, an independent rater also filled in the scale. The inter-rater-agreement for the four scales (Cohen’s Kappa Scores) separately was .75, .43, .72 and .57, respectively. Because the total scores of both raters did not cover the same range, a Spearman correlation coefficient was calculated. This turned out to be high and significant (.97).

**Adjunct outcome measures**

Another measure for treatment effectiveness was derived from Goal Attainment Scaling, a method used to measure level of attainment of individual goals, originally introduced by Rockwood, Joyce, and Stolee (1997). In our version, Treatment Goal Attainment (TGA), patients in both groups had to determine three personal goals they wanted to accomplish by means of the training. This took place after ten training sessions, after patients had been given the opportunity to gain some insight into their strengths and weaknesses. After treatment, patients were asked to indicate on a 5-point scale (1 = not at all, 5 = entirely) to what extent they had attained each goal; the total score thus ranged from 3 to 15. At follow-up, patients were asked to rate this again.

To measure EF in a complex task, a newly designed test was administered at follow-up only. The Executive Secretarial Task (EST; Lamberts, Evans, & Spikman, 2009; Spikman, Hol-Steegstra, Rietberg, Vos, Boelen, & Lamberts, 2007) is an ecological EF task, in which a job assessment procedure is simulated. This 3-hr task is comparable to the Multiple Errands Tasks (Shallice & Burgess, 1991) or the Hotel Task (Manly, Hawkins, Evans, Woldt, & Robertson, 2002) in that it requires the organization, initiation and prioritisation of multiple tasks over a longer time span than usual, while dealing with delayed intentions, interruptions and deadlines. The task yields three scores, together forming a Total Score: Initiative; reflecting all the actions the subject has initiated without being told so, Prospective; reflecting all the actions that were correctly carried out in a later stage, and Executive, reflecting all the actions that were correctly carried out at all. The EST was administered only at follow-up. This warranted a ‘pure’ indication of executive functioning without contamination of previous test knowledge or experience.

**Questionnaires and observation lists**

The presence of executive symptoms in everyday life was investigated by means of the Dysexecutive Questionnaire (DEX; Burgess et al., 1996), with a patient, proxy and therapist version.

The Executive Observation Scale (EOS; based on Pollens, McBratnie, & Burton, 1988) consists of eight items covering the EF aspects distinguished by Ylvisaker. Items are rated by the therapist on a scale from 1 (complete inability) to 4 (complete independence) with a total score ranging from 8 to 32. Quality of life was measured with the Quality of Life after Brain Injury (QOLIBRI (early version); Von Steinbüchel, Petersen, & Bullinger, 2005) filled in by the patient. The QOLIBRI consists of two parts; a satisfaction scale (higher score indicates more satisfaction) and a burden scale (higher score indicates higher burden).

At follow-up, patients were requested to rate their levels of satisfaction with treatment (results) on a 5-point Treatment Satisfaction Scale (TSS), ranging from score 1 (not satisfied) to score 5 (very satisfied).

**Neuropsychological measures of executive and cognitive functioning**

In addition to the mentioned behavioural measures, the following EF tests were administered. The Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996); all six subtests, resulting in a Standard Age Score. The Trail Making Test (ratio B vs. A, TMT B/A), Stroop Test (Stroop, 1935) (ratio time part three vs. time part two, Stroop 3/2); Tower of London (Shallice, 1982), (TOL, number correct).

To control for possible effects of the Cogpack training on memory, the 15 Words Test (Dutch version of RAVLT (Deelman, Brouwer, van Zomeren, &
After inclusion, patients were randomly assigned to one of the treatment conditions. Table 1 shows that at baseline no significant differences between the experimental and control patient groups were present on several relevant measures, such as indications of cognitive functioning (IQ, memory) and EF (Stroop 3/2, TMT B/A, TOL, BADS). Nor were there any differences on questionnaires measuring dysexecutive complaints (DEX), executive functioning observed during task performance by therapists (EOS), the extent to which previous roles were resumed (RRL) and the satisfaction part of the Quality of Life questionnaire (QOLIBRI). The only significant difference regarded the QOLIBRI Burden Score: before treatment experimental patients experienced a higher burden due to brain injury than control patients.

Effects of treatment
Table 2 shows the results of repeated measures analyses on the primary outcome measure, the RRL. After treatment (T1) both groups had resumed their previous roles significantly more than before treatment, but the experimental patients to an even larger extent. The same result was found for immediate (Memory IR) and delayed recall score (Memory DR), was administered at baseline and T1.

Statistical analyses
T-tests were used to compare patients’ scores on relevant measures at baseline with those of healthy controls, as well as to verify whether both patient groups differed at baseline.

Treatment effects were analyzed as follows. The primary outcome measure, the RRL, as well as those adjunct outcome measures for which a pre- and post measurement was available, were analyzed using repeated measures analyses (GLM repeated measures, SPSS 16.0). Because data loss due to missing values was undesirable, all test-measures were analyzed separately in a univariate design, in three series. First, it was tested whether scores at T1 differed from baseline, visible in a time-effect, and whether improvement over time differed between experimental and control patients, reflected in an interaction effect. Similarly, scores at T2 were compared to baseline performance, as well as to performance on T1. To overcome the probability of spurious results, Bonferroni Holm corrections were applied.

Because there were no baseline measures for the TGA and the EST, results at T1 and T2 were compared using t-tests. EST results of both patient groups were also compared with those of a healthy control group. Finally, the skewed scores on the TSS were compared using a non-parametric Mann-Whitney U test.

Because we considered executive performance at follow-up the most important indication of treatment success, a two-way multivariate analysis of variance (MANOVA) was performed on the measures that represented various relevant facets of executive functioning in daily life, namely the EST, TGA, and RRL.

Results
To verify whether the patients had impaired EF and thus fulfilled inclusion criteria, their BADS and DEX scores were compared to those of a healthy control group. Table 1 shows that both groups were comparable with respect to educational level. However, healthy controls were slightly older, but as age is known to influence executive functioning negatively, this was not advantageous. On the BADS, patients showed significantly worse EF than healthy controls. Also, patients had higher scores on the three DEX-measures evidencing more executive problems in daily life.

After inclusion, patients were randomly assigned to one of the treatment conditions. Table 1 shows that at baseline no significant differences between the experimental and control patient groups were present on several relevant measures, such as indications of cognitive functioning (IQ, memory) and EF (Stroop 3/2, TMT B/A, TOL, BADS). Nor were there any differences on questionnaires measuring dysexecutive complaints (DEX), executive functioning observed during task performance by therapists (EOS), the extent to which previous roles were resumed (RRL) and the satisfaction part of the Quality of Life questionnaire (QOLIBRI). The only significant difference regarded the QOLIBRI Burden Score: before treatment experimental patients experienced a higher burden due to brain injury than control patients.

Effects of treatment
Table 2 shows the results of repeated measures analyses on the primary outcome measure, the RRL. After treatment (T1) both groups had resumed their previous roles significantly more than before treatment, but the experimental patients to an even larger extent. The same result was found for

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Means (SD) of the primary outcome measure Role Resumption List (RRL) at T0, T1 and T2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group (n = 37, at T2 n = 36)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>RRL T0</td>
<td>7.8 (3.6)</td>
</tr>
<tr>
<td>RRL T1</td>
<td>7.2 (3.3)</td>
</tr>
<tr>
<td>RRL T2</td>
<td>7.4 (3.2)</td>
</tr>
<tr>
<td>M T1-T0</td>
<td>-0.1</td>
</tr>
<tr>
<td>M T2-T0</td>
<td>-0.1</td>
</tr>
<tr>
<td>M T2-T1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Note: Means of the individual difference scores (T1-T0, T2-T0, T2-T1) and repeated measures analyses on the test-scores at T1 compared to T0, T2 compared to T0, and T1 compared to T2. *p < .05; **p < .01; ***p < .001.
Table 3 shows the results of repeated measures analyses on the adjunct outcome measures at T1 compared to baseline. From post-treatment to follow-up, only the experimental group showed further improvement over time. Additional univariate analyses of variance were performed at T1 and T2 in order to find out whether there were differences between treatment centers; both analyses showed that this was not the case (T1: $F(1,17) p = .34$; T2: $F(1,88), p = .124$).

Table 4 shows the results of repeated measures analyses at T2 compared to T1. Similar to T1, both groups showed significantly larger improvements in the experimental group than the control group. This time the effect for the DEX-patient and the DEX-proxy showed a time effect. The time effect for the DEX-therapist was found in both groups. Both groups also showed improvement in the EOS, with a significant larger effect for the experimental group. On the standard EF tests (Stroop 3/2, TMT B/A, TOL, and BADS), no interaction was found indicating differential treatment effects. The TOL and the BADS showed a time effect, but the effect was similar for both groups. The BADS showed a significant larger effect for the experimental group. The TOL showed a significant larger effect for the experimental group. No effects were found on the QOLIBRI scales.

Table 5 shows the means and (SD) and means of individual difference scores (T1-T0) of the adjunct outcome measures at T1 for both experimental and control groups. The TSS showed no difference, indicating that both experimental and control groups were equally satisfied with the treatment results.
Table 4: Means (and SD) and means of individual difference scores (T2-T0) of the adjunct outcome measures at T2 for control and experimental groups

<table>
<thead>
<tr>
<th></th>
<th>Control group (n = 36)</th>
<th>Experimental group (n = 35)</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M T2-T0</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Stroop 3/2</td>
<td>1.6 (0.3)</td>
<td>-0.1</td>
<td>1.5 (0.2)</td>
</tr>
<tr>
<td>TMT- B/A</td>
<td>2.2 (1.3)</td>
<td>-0.1</td>
<td>2.1 (0.6)</td>
</tr>
<tr>
<td>TOL nr correct</td>
<td>11.0 (1.1)</td>
<td>0.2</td>
<td>11.0 (1.4)</td>
</tr>
<tr>
<td>BADS</td>
<td>94.6 (13.2)</td>
<td>8.8</td>
<td>101.8 (12.5)</td>
</tr>
<tr>
<td>DEX patient</td>
<td>255 (44.6)</td>
<td>-6.1</td>
<td>263 (15.8)</td>
</tr>
<tr>
<td>DEX proxy</td>
<td>285 (16.6)</td>
<td>-2.2</td>
<td>265 (16.7)</td>
</tr>
<tr>
<td>DEX therapist</td>
<td>289 (35.3)</td>
<td>-6.8</td>
<td>242 (11.2)</td>
</tr>
<tr>
<td>QOLIBRI satisfac</td>
<td>136.4 (35.4)</td>
<td>12.7</td>
<td>126.7 (32.4)</td>
</tr>
<tr>
<td>QOLIBRI burden</td>
<td>44.1 (12.0)</td>
<td>-4.2</td>
<td>51.7 (18.1)</td>
</tr>
<tr>
<td>EOS</td>
<td>223 (4.2)</td>
<td>1.6</td>
<td>257 (35.5)</td>
</tr>
</tbody>
</table>

Note: Repeated measures analyses on the test scores at T2 compared to T0. TMT = Trail Making Test; TOL = Tower of London; BADS = Behavioural Assessment of the Dysexecutive Syndrome; DEX = Dysexecutive Questionnaire; EOS = Executive Observation Scale; n.s. = not significant.

* Significant p value < Bonferroni Holm corrected alpha.

Table 5: Means (and SD) of the TGA at T1 and T2, and for the TSS at T2 for the control and experimental group

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Experimental group</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M T1-T2</td>
<td>M (SD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGA T1</td>
<td>8.2 (1.2)</td>
<td>4.4</td>
<td>10.7 (2.0)</td>
</tr>
<tr>
<td>TGA T2</td>
<td>8.9 (2.0)</td>
<td>2.9</td>
<td>10.6 (2.0)</td>
</tr>
<tr>
<td>TSS</td>
<td>3.8 (0.9)</td>
<td>-0.4</td>
<td>3.5 (0.9)</td>
</tr>
</tbody>
</table>

Note: Two-tailed results of t-tests. TGA = Treatment Goal Attainment; TSS = Treatment Satisfaction Scale; n.s. = not significant.

* p < .05; ** p < .01; *** p < .001.

Table 6: Means (and SD) of control and experimental groups and healthy controls on the total score and the three subscores of the EST

<table>
<thead>
<tr>
<th>EST (group)</th>
<th>CG (n = 34)</th>
<th>EG (n = 32)</th>
<th>HC (n = 57)</th>
<th>CG vs EG</th>
<th>CG vs HC</th>
<th>EG vs HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (0-45)</td>
<td>28.3 (10.2)</td>
<td>34.1 (8.2)</td>
<td>37.4 (5.6)</td>
<td>2.5***</td>
<td>-4.8***</td>
<td>-2.3*</td>
</tr>
<tr>
<td>Initiative (0-13)</td>
<td>9.5 (3.1)</td>
<td>10.7 (2.5)</td>
<td>11.5 (1.9)</td>
<td>1.6 n.s.</td>
<td>-3.3**</td>
<td>-1.8 n.s.</td>
</tr>
<tr>
<td>Prospective (0-8)</td>
<td>5.2 (2.2)</td>
<td>6.8 (1.5)</td>
<td>6.7 (1.4)</td>
<td>3.4**</td>
<td>-3.7***</td>
<td>0.3 n.s.</td>
</tr>
<tr>
<td>Executive (0-24)</td>
<td>13.6 (6.2)</td>
<td>16.6 (5.9)</td>
<td>19.1 (3.9)</td>
<td>2.1*</td>
<td>-4.7***</td>
<td>-2.2*</td>
</tr>
</tbody>
</table>

Note: Two-tailed results of t-tests. EST = Executive Secretarial Task; CG = Control Group; EG = Experimental Group; HC = Healthy Controls.

* p < .05; ** p < .01; *** p < .001.
indications of daily life functioning showed significant differences between both groups after treatment; the TGA, reflecting the ability to set and accomplish realistic goals, and the EST, reflecting the ability to plan, organize and regulate a series of real-life tasks. On the TGA, differences were evident after treatment as well as at follow-up, indicating that the patients of the experimental group had attained their goals to a larger extent than the control patients. In the EST, only administered at follow-up, the experimental group performed significantly better on all scores, except for the Initiative subscore.

The RRL was the only measure that was administered blindly at every measurement. For different reasons both the TGA and the EST could not be administered at baseline. The TGA had to be an indication of the ability to set realistic treatment goals, for which participants had to gain insight into their strengths and weaknesses in the first part of the treatment. In both treatment conditions goals were therefore set after the first 10 sessions. The EST is a complex problem-solving test, which is not appropriate for use in test-retest conditions. When EF tests are administered repeatedly, crucial aspects of novelty and problem-solving become biased by learning effects (for instance, retaining solutions and strategies in memory), which will influence performance a second time. Hence, these effects can not be disentangled from the pure executive problem-solving components of test performance. Therefore, the BADS was not considered appropriate as a primary outcome measure in the present study. As expected, no treatment effects were found on the BADS; compared with baseline, both groups had improved to the same extent at T1 and T2, which we interpret as a test-retest effect. Previous evidence for a considerable test-retest effect on the BADS was found by Jelicic, Henquet, Derix, and Jolles (2001).

Although significant group differences were found on the TGA and EST, the lack of pre-treatment measures implies that these differences cannot be ascribed as strongly to treatment effects as in the case of the RRL. Nevertheless, there are robust indications that these differences are due to therapy. After all, both groups were carefully randomized (see Table 1) and they did not differ significantly with respect to their initial executive and cognitive capacities, nor with regard to biographical variables at baseline.

The significant difference found with a MANOVA performed on the combination of the three effect measures (RRL, TGA and EST) at follow-up, indicated even more strongly that treatment effects were still present and clinically substantial at follow-up.

In our multifaceted treatment, several elements of proven treatment methods had been incorporated, namely Goal Management Training (Levine Table 7 shows the results of a MANOVA on indications of daily life executive functioning (EST, RRL and TGA) at T2. A significant treatment effect on the combination of these three measures was found. Univariate analyses showed that this could be mainly attributed to the RRL and the TGA.

**Table 7** Results of a multivariate analysis (MANOVA) at T2 on the EST, the TGA, and the RRL for the control and experimental groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>F(df)</th>
<th>Sign.</th>
<th>Variable</th>
<th>F(df)</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7.9 (3,62)</td>
<td>p &lt; .001</td>
<td>EST</td>
<td>2.5 (1,65)</td>
<td>n.s.</td>
</tr>
<tr>
<td>TGA</td>
<td>5.1 (1,65)</td>
<td>p &lt; .05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>3.0 (3,62)</td>
<td>p &lt; .05</td>
<td>RRL</td>
<td>5.8 (1,65)</td>
<td>p &lt; .05</td>
</tr>
</tbody>
</table>

Note. EST = Executive Secretarial Task; TGA = Treatment Goal Attainment; RRL = Role Resumption List; n.s. = not significant.

**Discussion**

The results of this study were two-fold. Those measures that were considered indications of daily life executive functioning showed that a multifaceted treatment for executive dysfunction after ABI led to a significant treatment effect for the experimental group compared to the control group, lasting at least 6 months post-treatment. However, other measures regarding well-being and subjective complaints showed similar improvements for both groups, whereas conventional cognitive and executive tests showed only time effects or no effects at all.

The improvement in daily life executive functioning was evidenced on the primary outcome measure, the RRL, indicating the ability to resume previous roles in work, social relations, leisure activities, and mobility and therefore of functioning at social participation level. Our expectation that these effects on daily life EF would be present after treatment and would subsequently remain stable for a substantial period was entirely met. The significant differences between both groups after treatment were even larger at follow-up, indicating that the experimental group had increased their wished-for roles in daily life even further, whereas the control group remained at the same activity level throughout. In addition, two other
and proxies experienced and observed less dysexecutive problems after lower scores on the DEX-patient and -proxy at T1 and T2 indicate that patients these patients appreciated the training and its effects on their lives. The throughout an intensive, laborious treatment. Results on the TSS show that our experimental patient group was able to adhere to and remain motivated of strategic behaviour and the ability to benefit from treatment. Nevertheless, 1991). After all, dysexecutive problems are known to hamper the acquisition dysexecutive problems, which is not always taken for granted (Alderman, 1991). After all, dysexecutive problems are known to hamper the acquisition of strategic behaviour and the ability to benefit from treatment. Nevertheless, our experimental patient group was able to adhere to and remain motivated throughout an intensive, laborious treatment. Results on the TSS show that these patients appreciated the training and its effects on their lives. The lower scores on the DEX-patient and -proxy at T1 and T2 indicate that patients and proxies experienced and observed less dysexecutive problems after treatment. However, this was also true for the control patients. With respect to quality of life, both groups also showed the same pattern of results. This result was surprising, because we feared that patients’ motivation to perform the long and energy sapping Cogpack training would fade or that they would sense that this treatment would not be effective. On the contrary, the majority of patients were very enthusiastic about Cogpack as it provided immediate feedback on performance, so that patients could monitor their improvement over time. This suggests that Cogpack influenced patients’ sense of self-efficacy positively, exactly as the experimental training did. Apparently, this has led to higher levels of activity in control patients, reflected by increased social participation, although not to the same extent as in patients of the experimental group. In addition, the control training also ameliorated trainees’ executive abilities, rated by therapists on the DEX-therapist and the EOS.

As expected, there was no indication that either of the treatments had significant effects on cognitive or executive functioning as measured with neuropsychological tests, especially on the Stroop, TMT, and TOL as well as the Memory test. This lack of effects on common executive tests indicates how difficult it remains to assess executive functions in daily life, and especially changes in daily executive functioning, with conventional neuropsychological tests. Additional research will be needed in the future to disentangle the relation between executive tests and daily life. The lack of effects on neuropsychological tests (including memory tests) also shows that Cogpack training did not live up to its promise of improving basic cognitive functions.

To summarize, control patients’ levels of satisfaction and subjective well-being were equal to those of the experimental group, although the latter group performed significantly better on measures that pertain to daily life executive functioning. This indicated that significant treatment effects can be accomplished by a multifaceted treatment, if it is aimed at improving activity and social participation and tailored to the individual patient. Moreover, these effects last for a substantial period after ending the treatment.

Acknowledgements

This study was supported by The Dutch Organization of Health Research and Development (ZON-MW), Rehabilitation Research Program (Grant No. 1435.0009)
The authors are grateful to all the participants and their families, as well as to the healthcare institutions and their neuropsychologists that participated in this study: Revalidatiecentrum Amsterdam (Amsterdam), Revalidatie Friesland (Beestkerzwaag and Leeuwarden), Academisch Revalidatiecentrum Beatrixoord (Haren), Revalidatiecentrum Helios (Wijk aan Zee), Revalidatiecentrum Groot Kimmendaal (Arnhem), Sint Maartenskliniek (Nijmegen), Revalidatiecentrum De Vogellanden (Zwolle) and University Medical Center Groningen (Groningen). We thank Anneke Hol, Sylvie Vos, and Hennerieke Rietberg for their assistance with the data collection.

References


Predictors of treatment outcome in a successful experimental treatment of executive dysfunction

In revision
D.H.E. Boelen, L. Fasotti, A.C.M. Rietveld, J.M. Oosterman, J.M. Spikman
Abstract

The aim of this short report was to describe two additional aspects of a recently published study on the successful treatment of executive dysfunction in daily life. This so-called multifaceted treatment was designed and applied to a group of patients with acquired brain injury. First, a total effect size for the treatment was calculated. Second, it was investigated whether individual differences in the degree of improvement of the experimental group were related to differences in scores on three types of predictor variables. The results showed that the total effect size of the multifaceted treatment was substantial. Moreover, correlations between predictor variables at baseline and three outcome measures were largely non-significant. Only IQ and the Tower of London were predictive of success, suggesting that the multifaceted treatment is applicable to a large variety of patients with executive impairments.

Introduction

Executive dysfunction (ED) is a common consequence of neurological disorder and is mostly investigated in patients with traumatic brain injury and stroke. ED is defined as the inability to adapt to new, non-routine situations and to plan and organise activities in a goal-directed manner. (Duncan, 1996; Levine et al., 2000). The last two decades there have been many attempts to treat executive disorders. However, the term executive functions is an ‘umbrella term’ which encompasses a broad range of functions. As a result of this diversity and the complexity of executive functions, the efforts to remediate them have been equally diverse. The starting point for some of these treatments was the rehabilitation of cognitive aspects of executive impairments, as exemplified in the compensation for problems with goal setting (e.g., Levine et al., 2000) and problem-solving (e.g., Rath, Simon, Langenbahn, Sherr, & Diller, 2003; von Cramon & Matthes-von Cramon, 1994). Other treatments have focussed more on the behavioural aspects of ED (e.g., Alderman, Fry, & Youngson, 1995).

Just as there exists a variety of treatments, several designs have been used to investigate treatment effects. Experimental designs varied from single case designs (e.g., Burke, Zencius, Wesolowski, & Doubleday, 1991) to controlled group studies (e.g., Manly, Hawkins, Evans, Woldt, & Robertson, 2002) and randomized controlled designs (e.g., Wilson, Emstle, Quirk, & Evans, 2001). The results of most of the intervention studies seem encouraging, in that (aspects of) executive dysfunction can be ameliorated. However, the verification of treatment effectiveness is often compromised by restrictions in the used experimental designs (Boelen, Spikman, & Fasotti, 2011). For example, group studies may lack important ingredients such as a control group, or the randomization of subjects may be lacking.

A randomized controlled trial (RCT) is considered the most reliable way to evaluate treatment effects in group studies. Double blind RCTs are rigorous in determining whether a cause-effect relation exists between a treatment and its outcome. They involve random allocation of subjects to different treatments, which reduces selection bias. Random assignment also ensures that differences between treatment groups behave like the differences between random samples from a whole population. This in turn, warrants a higher external validity. Moreover, blinding (i.e., masking both subjects and investigators to the treatment) minimizes selection bias as well as ascertainment bias (i.e., bias due to systematic distortion in the outcome measures) (Rawlins, 2008).

Recently, Spikman, Boelen, Lamberts, Brouwer, and Fasotti (2010) published a study on the effects of a treatment program for ED after brain
injury. The treatment was aimed at improving several aspects of executive functioning and was therefore called “multifaceted”. The patients had acquired brain injury of various aetiologies and were in the chronic stage of recovery. Treatment efficacy was evaluated by means of a RCT in which the experimental group received the multifaceted treatment for executive dysfunction and the control group underwent a computerized cognitive function training. Following the results of Spikman et al. (2010), three daily life executive outcome measures showed improvement after treatment. Improvement was evidenced in the primary outcome measure, the Role Resumption List (RRL; Spikman, Brand, & Brouwer, 2003), indicating the ability to resume previous roles in work, social relations, leisure activities, and mobility. A significant difference between the experimental and the control group was also found in two other executive indicators (for a description see below). Because the Spikman et al. (2010) study was a RCT, any significant differences between groups in the outcome can be attributed to the intervention and not to some other, unidentified factor. Thus, it is safe to conclude that the significant improvement on the RRL and the other two outcome measures in the experimental group of that study can be attributed to the multifaceted treatment of ED.

Although experimental and control groups differed on these measures, the total effect of the multifaceted treatment has so far not been established (Spikman et al., 2010). Thus, in this short notice, a total effect size, combining the three outcome measures, will be computed. Such an effect size conveys the average difference between two groups without any discussion of the variability within the groups.

On the other hand, investigating within-group variability is a relevant additional question. More specific: did all individuals within the successful (experimental) group in the Spikman et al. (2010) study benefit from the treatment to the same degree? Analyses of the effects of an intervention in subgroups of patients can be important to establish whether different types of patients respond differently (Rawlings, 2008). In other words, did background and baseline variables influence individual outcome scores and/or were treatment benefits confined to certain subgroups of patients? The group that received the multifaceted treatment for ED was more successful on the three outcome measures at follow-up than the control group that received a standard cognitive function training. Nevertheless, not every patient in this successful group may have improved to the same degree.

The main aim of this short notice is to report whether individual differences in the degree of improvement in the experimental group of the Spikman et al. (2010) study relate to differences in scores on three types of predictor variables: background variables, neuropsychological test scores and behavioural measures. Background variables pertain to patient characteristics such as age, IQ and aetiology. Executive impairment was assessed with neuropsychological tests and questionnaires at baseline.

Method

Subjects
Seventy-five patients participated in the study of Spikman et al. (2010) that was set up as a RCT. There were 38 patients in the experimental group who received the multifaceted treatment for executive dysfunction. The control group consisted of 37 patients and received an individually administered computerized cognitive function training. Patients were included in the study on the basis of complaints of executive problems in daily life as reported by themselves and/or observed by proxies. The Dysexecutive Questionnaire (DEX; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) was used to quantify executive complaints. Inclusion was also based on BADS scores and IQ scores (see Spikman et al., 2010). Mean IQ of the experimental group was 116.1, SD 16.8. All the patients were in the chronic stage of recovery.

Data were obtained according to the ethical regulations of the participating institutions, in compliance with the Helsinki Declaration.

Predictor variables
The predictor variables consisted of background variables, test scores and questionnaires. The background variables were age, IQ and aetiology. Test scores at baseline were BADS total standard score (Behavioural Assessment of the Dysexecutive Syndrome; Wilson et al., 1996) and TOL total correct (Tower of London; Shallice, 1982). The total scores of two pre-treatment questionnaires were used: the DEX-patient score and the PCRS-patient score (Patient Competency Rating Scale, PCRS; Prigatano et al., 1986). A full description of these tests and questionnaires can be found in Boelen, Spikman, Rietveld, and Fasotti (2009).

Dependent variables
The primary outcome measure was the Role Resumption List (RRL; Spikman et al., 2003). Based on a structured interview with the patient, the RRL assesses changes in the amount and quality of activities compared to pre-morbid levels in four daily life domains (vocational functioning, social interaction, leisure activities, and mobility). These domains were rated on a
five-point scale (0 = no change, 4 = severe loss of independence), with a total score ranging from 0 to 16. A higher score indicated a greater loss of roles and independence. The RRL rating procedure was double-blind: The RRL was administered at baseline (T0), after treatment (T1) and at follow-up (T2). An adjunct outcome measure was Treatment Goal Attainment (TGA), a derivate from Goal Attainment Scaling. Patients were asked to formulate three goals during the treatment, which they were asked to rate on a 5-point scale at T1 and T2. Another adjunct outcome measure, the Executive Secretarial Task (EST; Lamberts, Evans, & Spikman, 2010) was administered only once (T2) to obtain a ‘pure’ indication of executive functioning after treatment, without contamination of previous test experience or knowledge. In all three cases rating procedures were double-blind: independent raters as well as patients were blind for treatment allocation.

Analyses
In addition to the results presented by Spikman et al. (2010) a composite outcome measure (Total Executive Outcome Score, TEOS) comprising the standardized (Z-)values of EST, TGA and RRL at follow-up was computed. For standardization the RRL scores were recoded, so that a high score indicated good performance. A $t$-test was used to test whether the experimental and control group differed on the TEOS. An effect size (Cohen’s $d$) was calculated using the difference between the mean scores of both groups divided by the pooled standard deviation. According to Cohen (1988), an effect of 0.2 is small, 0.5 is medium, and 0.8 is large.

With respect to the analysis of individual differences in performance on three outcome measures, one outcome measure was available both at baseline and follow-up (RRL). For this variable a difference score (T2-T0) could be computed and used as a dependent variable (diffRRL). TGA and EST were also significantly better in the experimental group at follow-up, but there were no baseline scores for these variables. For these two variables the scores at T2 were used in the analysis.

To investigate whether individual differences in the degree of improvement on the diffRRL, EST and TGA relate to differences in scores at baseline measures within the successful experimental group, correlations were calculated between the three outcome measures and background variables (IQ, age), executive test results (BADS, TOL) and the scores on executive questionnaires (DEXpt, PCRSpt). Pearson correlations were used for the diffRRL and TGA. Skewed data on the EST required non-parametric Spearman correlations.

In addition, there was one non-continuous background variable: aetiology (left hemisphere stroke, right hemisphere stroke, subcortical stroke, traumatic brain injury, otherwise). With this variable correlations could not be computed and therefore three separate Kruskal-Wallis tests were used, one for each outcome measure.

Results
A two-tailed $t$-test showed that the patients in the experimental group ($N=31$) scored significantly higher on the TEOS (composite outcome measure) than the patients in the control group ($N=35$) ($t = -3.8; df = 64; p < .001$). The effect size for this difference was 0.86. According to Cohen’s criteria this is a large and clinically relevant therapy effect of the experimental treatment on daily life executive functioning.

Correlations between the three outcome measures and background variables (IQ, age), executive test results (BADS, TOL) and the scores on executive questionnaires (DEXpt, PCRSpt) are presented in Table 1.

Table 1 Correlations between predictor variables and outcome measures in the experimental group

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>DiffRRL Pearson (N = 33)</th>
<th>EST Spearman (N = 31)</th>
<th>TGA Pearson (N = 35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.117 n.s.</td>
<td>-.176 n.s.</td>
<td>-.137 n.s.</td>
</tr>
<tr>
<td>IQ</td>
<td>.387*</td>
<td>.504**</td>
<td>.119 n.s.</td>
</tr>
<tr>
<td>BADS</td>
<td>.307 n.s.</td>
<td>.172 n.s.</td>
<td>-.055 n.s.</td>
</tr>
<tr>
<td>TOL</td>
<td>-.376*</td>
<td>.367*</td>
<td></td>
</tr>
<tr>
<td>DEXpt</td>
<td>.260 n.s.</td>
<td>.114 n.s.</td>
<td>-.288 n.s.</td>
</tr>
<tr>
<td>PCRSpt</td>
<td>.039 n.s.</td>
<td>-.094 n.s.</td>
<td>.255 n.s.</td>
</tr>
</tbody>
</table>

Note: BADS = Behavioural Assessment of the Dysexecutive Syndrome; TOL = Tower of London; DEX = Dysexecutive Questionnaire; PCRS = Patient Competency Rating Scale.
* p<.05, **p<.01
Discussion

A RCT on the effectiveness of a multifaceted treatment for ED after acquired brain injury was recently performed (Spikman et al., 2010). Seventy-five patients were randomly assigned to either an experimental treatment or a placebo treatment. The 38 patients in the experimental group were significantly more successful on three outcome measures than the 37 patients in the control group. The current short report provides an effect size of the difference between the two groups. The main aim however, was to examine whether differences in improvement within the experimental group on the three outcome measures were related to baseline differences between subjects in this group.

A large effect size was found on the TEOS, the total executive outcome score composed of the three outcome measures RRL, TGA and EST at follow-up. This indicates that treatment effects were still present and substantial at follow-up. These results show that the dysexecutive syndrome is not only treatable in this particular group of patients, but that high effect sizes can be attained on measures that assess daily executive functioning. This refutes the idea that executive problems are difficult or even impossible to treat. Previous treatments have mainly been aimed at single aspects of ED; the present study proves that it is possible to treat ED more comprehensively, with considerable effects on daily life executive performance.

The main topic of this brief report was to investigate whether individual differences in the degree of improvement in the experimental group of the Spikman et al. study (2010) could be predicted by several kinds of variables. Patients with a higher IQ had lower scores on the RRL at baseline, which means that they had better role resumption at the beginning of the treatment than patients with a lower IQ. Thus, at T0 and T2 a higher IQ was associated with lower scores on the RRL, indicating that patients with a higher IQ had resumed their roles more fully than patients with a lower IQ already before treatment (and this had remained so six months after treatment). Thus, although the improvements in role resumption of patients with a higher IQ were smaller, these patients had resumed their previous (i.e., premorbid) roles to a higher degree before treatment (lower RRL-scores) than patients with a lower IQ. This indicates that, for patients with a higher IQ there was less room for improvement in role resumption compared with patients with a lower IQ. IQ also positively correlated with the EST, showing that at follow-up patients with a higher IQ performed significantly better on this complex and lengthy executive test.

The TOL correlated negatively with diffRRL, indicating that a higher score on the TOL is accompanied by a lower score on diffRRL, i.e., better role resumption. The TOL was also positively correlated with TGA, indicating that higher TOL-scores were accompanied by a significantly better attainment of treatment goals.

The primary variable diffRRL correlated significantly with IQ (r = .387, p < .05) (N=33) and TOL (r = -.376, p < .05) (N=33).

The EST correlated significantly with IQ (r = .504, p < .01; N=31), whereas TGA correlated significantly with TOL (r = .367, p < .05; N=33). Thus, of the background variables only IQ correlated significantly with two of the outcome measures (diffRRL and EST). Neither questionnaire correlated significantly with any of the three outcome measures. Regarding the neuropsychological tests only the TOL correlated with effect measures (diffRRL and TGA). It should be noted however, that these correlations were moderate.

Kruskall-Wallis tests showed that aetiology did not influence progress (diffRRL) or performances (EST, TGA).

IQ correlated positively with diffRRL (r = .387, p < .05), indicating that a higher IQ was associated with higher diffRRL scores. Higher scores on the diffRRL point to less improvement in role resumption. However, the correlation between IQ and RRL is negative (r = -.366, p < .05; N=33) at T2 as well as at T0 (r = -.622, p < .01; N=36). Thus, at T0 and T2 a higher IQ was associated with lower scores on the RRL, indicating that patients with a higher IQ had resumed their roles more fully than patients with a lower IQ already before treatment (and this had remained so six months after treatment). Thus, although the improvements in role resumption of patients with a higher IQ were smaller, these patients had resumed their previous (i.e., premorbid) roles to a higher degree before treatment (lower RRL-scores) than patients with a lower IQ. This indicates that, for patients with a higher IQ there was less room for improvement in role resumption compared with patients with a lower IQ. IQ also positively correlated with the EST, showing that at follow-up patients with a higher IQ performed significantly better on this complex and lengthy executive test.

The TOL correlated negatively with diffRRL, indicating that a higher score on the TOL is accompanied by a lower score on diffRRL, i.e., better role resumption. The TOL was also positively correlated with TGA, indicating that higher TOL-scores were accompanied by a significantly better attainment of treatment goals.
general intelligence are closely related to Spearman’s g and the g-factor is a reflection of frontal lobe function (Duncan, Ensmie, Williams, Johnson, & Freer, 1996).

Among the neuropsychological tests, the TOL correlated significantly with two of the outcome measures (diffRRL and TGA). Higher scores on the TOL were accompanied by a decrease in the score on the diffRRL (i.e., indicating resumption of premorbid roles), and a greater attainment of treatment goals. Apparently, the TOL is a significant predictor of the success of the multifaceted treatment. The TOL might be an effective baseline assessment tool that can be used to include patients in the treatment.

Neither questionnaire correlated with any of the outcome measures. Thus, the amount of subjective complaints was not predictive of the degree of improvement at follow-up. Hence, it seems likely that the treatment can be administered irrespective of the amount and severity of complaints of executive impairments that patients report. This implicates that patients with either milder or more severe subjective complaints of ED might benefit from the treatment.

To summarize, this short report aimed to analyze variability within the experimental group, by relating individual outcome scores to baseline variables. The results showed that especially patients with a low IQ benefited from the treatment. The TOL seems to be a useful tool to include patients in the treatment. The clinical message derived from the current findings is that a multifaceted treatment for ED is effective for a broad range of patients with a dysexecutive syndrome.

References


Spikman, J.M., Brand, B., & Brouwer, W.H. (2003). Roolhervattingslijst [Role Resumption List]. Internal Publication of the University Medical Centre Groningen, the Netherlands. Groningen: Groningen University Medical Centre.


General discussion and conclusions
General discussion and conclusions

For a long time executive disorders have been mainly studied in patients with lesions in the frontal lobes. Progressively, however, both clinicians and researchers have recognised the need for an adjustment of this idea. Several investigations have shown that not only patients with frontal lesions, but also patients with lesions in other brain regions that are connected with the frontal lobes, may suffer from executive impairments. The implication of these findings is that considerably more patients have executive impairments than originally thought, when numbers were solely based on patients with frontal brain damage. Therefore, a substantial number of brain-injured patients encountered in rehabilitation are in need of treatment for their executive problems.

The patients who participated in the studies described in this thesis experienced significant executive impairments in their personal, social and vocational lives. Nevertheless, their ability to benefit from rehabilitation treatment was not entirely threatened by the severity of their dysexecutive problems. The executive complaints of the examined subjects could not always be linked to lesions of prefrontal brain regions. Also, the aetiology of brain damage in these subjects was diverse (stroke, traumatic brain injury (TBI), tumours, etc.).

The first aim of this thesis was a diagnostic one, namely to examine whether the executive complaints of this mixed group of brain-injured patients could be validated using neuropsychological instruments and which of these instruments were the most sensitive ones to assess the patients’ complaints. Likewise, the script processing abilities of patients with executive problems were assessed with a script generation task as a measure of daily life planning skills. The second aim of this thesis was to investigate whether brain-injured patients with executive complaints referred for outpatient rehabilitation, could successfully be treated with a new rehabilitation program. This program was a multifaceted treatment, which aimed at several executive aspects simultaneously. As the concept of executive function is an umbrella term which encompasses a broad range of higher order capacities (such as planning, initiation, problem-solving and regulation) and executive dysfunction is an equally heterogeneous concept, a multifaceted approach was chosen in order to treat a comprehensive range of executive problems these patients suffer from.

The study addressing the first research question, the assessment of executive complaints with neuropsychological instruments (described in chapter 2),
suggests that neuropsychological tests as well as questionnaires are sensitive to the executive problems of the patients. These findings confirm that not only patients with prefrontal lesions, but also subjects with lesions in other brain areas, can have executive problems. Therefore, it is reasonable to suggest that brain-injured patients suffering from executive dysfunction in daily life outnumber patients with frontal lesions. The same study also showed that a specific combination of tests and questionnaires (DEX-self, VF, EDT, BADS KS, BADS MSET) was more sensitive in tapping executive complaints than either tests or questionnaires. Thus, it appears that patients with executive complaints can be reliably assessed using a short screening battery, instead of an extensive and often laborious neuropsychological assessment procedure. In today’s medical climate emphasizing cost-efficiency and pointing to difficulties of reimbursement for elaborate assessment procedures, such a short screening might prove to be an effective and cost-cutting way to filter-out patients who are eligible for rehabilitation.

However, several issues of the abovementioned study are open to debate. The most evident shortcoming in this exploratory assessment study was the lack of a control group of brain-injured patients without executive complaints. Due to this absence it is not possible to rule out that the results of the study are attributable to a more general effect of brain injury, i.e. that the neuropsychological tests and questionnaires are sensitive to several non-specific effects of brain damage (e.g. deficits in mental effort, attentional deficits), and not specifically to executive impairments. Future research into this matter should involve such a control group without executive complaints.

Another limitation of the study is the absence of a golden standard against which neuropsychological tests and questionnaires should be validated. As mentioned in chapter 2, a solution for this problem through an improvement of the assessment of executive problems in daily life presupposes in-depth interviewing of patients, their proxies and therapists. Information gathered from these interviews should be complemented with information from valid questionnaires. At the same time, qualitative and quantitative observations, either structured or semi-structured, should be collected from brain-injured subjects with and without executive problems and from healthy control subjects. This information could be correlated with imaging data of the distributed neural network underlying executive functioning. This collection of anamnestic and hetero-anamnestic information, supplemented with behavioural observations and neuroimaging evidence should be the key element against which neuropsychological data should be compared in order to investigate the internal and external validity of neuropsychological tests. Another essential contribution to the improvement of neuropsychological assessment of executive disorders should come from cognitive-neuropsychological scientific investigation. Well-specified cognitive-neuropsychological models, as found in reading and writing (Coltheart, Brunsdon, & Nickels, 2005), may be helpful in determining with more accuracy the essential processes implied in (daily) executive functioning, as recently done by Miyake, Friedman, Emerson, Witzki, & Howeter (2000), in order to serve as a guide to the construction of better validated tests, questionnaires and observation scales. Therefore, the further investigation of the fractionation of executive processes (see General Introduction in chapter 1) is of the utmost importance in the future, not only for research, but also for clinical purposes.

Besides assessment with clinical neuropsychological tests, patients’ executive abilities were also assessed with a script generation task as presented in chapter 3. The results of this scripts study showed that patients indeed have significant problems with script generation, which again corroborates that the patients included in the present study on the basis of subjective complaints of executive problems in daily life, show unequivocal difficulties in the planning of action sequences needed to achieve a goal. Script generation therefore proves to be a sensitive task to investigate such impairments. Moreover, script generation planning of everyday activities is assessed. Hence, it is a task with high external validity, congruent with the problems that patients encounter in their everyday lives. This is in contrast to most conventional tasks that are used to assess executive functioning. Similar to everyday executive problems, script generation tasks provide a form of adaptive problem-solving, i.e. they involve intentionality and priority-based decision-making in relationship to the demands of an external situation (Goldberg & Podell, 2000). They are open-ended tasks that lack the structure of the usual neuropsychological examination, so that they are suitable to expose executive deficits, that otherwise may remain unnoticed. A more practical suggestion is that script generation tasks could quite easily be transformed into a structured neuropsychological test of which several parallel versions can be made. Another particularly advantageous aspect of script tasks is that they are essentially verbal tasks, whereas many other executive tasks use visual material. Script tasks tap inner speech in problem-solving situations. They probably better uncover how subjects think and reason in complex problem-solving situations and how behaviour is guided by verbal self-instruction.

An issue that is up for discussion in the scripts study is that only script generation was examined without investigating script execution. Complex
problem-solving actually consists of two stages (e.g. Luria, 1973; Lezak, 1995). The first stage is planning in order to predetermine a course of actions leading to a stated goal. The second stage entails the execution and monitoring of the plan to a successful conclusion. Some studies have shown that patients with executive disorders indeed had problems with script generation, which were even more pronounced when they executed the scripts (Chevignard et al., 2000; Fortin, Godbout, & Braun, 2003). Patients made significantly more planning as well as execution errors (e.g. omissions, errors of context neglect, environmental adherence, commentaries) than control subjects. These studies have shown that the execution of script tasks better predict executive impairments than other cognitive tests. Moreover, script generation and execution may be a more ecologically valid approach to estimate the severity of executive disorders in daily life activities. The execution of a plan involves a behavioural adaptation requiring the maintenance of the plan over longer periods of time and in the face of competing alternatives. This is obvious in daily life situations, but not in conventional neuropsychological tests. Generation and execution of scripts might simulate these demands better than laboratory testing, making this task a more ecologically valid alternative to the majority of executive tests. Thus, in future studies that validate script generation tasks, the scripts should be executed as well.

Another implication that follows from the script generation study is that scripts can be used as treatment tools. Before training the execution of everyday tasks, explicit consideration should be given to the formulation of the actions needed to fulfil such tasks. Within Goal Management Training for example (Levine et al., 2000), the construction of goal-lists is a prerequisite for ensuing task execution. In other words, a subject has to ‘disambiguate’ a situation and constrain it by selecting adequate sub-goals out of many possible alternatives (Goldberg & Podell, 1999). Through this process of constraining and selecting alternatives a script unfolds that guides behaviour toward goal attainment. Our scripts study has shown that patients were impaired in generating the exact actions, a deficit that might be underlying the problems that they have in executing everyday activities.

With regard to the second aim of this thesis, it can be concluded that it is indeed possible to effectively treat patients with executive disorders. The review study presented in chapter 4 showed that there is a panoply of possible treatments for executive impairments. However, each of the reviewed studies targeted different executive aspects. For clinicians it may therefore be hard to choose from this myriad of treatments, which intervention fits their patients best. In the review study we concluded that hardly any of the treatments covered more than one or two executive aspects, while patients usually present with various executive symptoms. The review study also showed that adequate research designs were not always used in studies of effectiveness of executive interventions. When the internal validity of a study is not warranted, valid conclusions can not be drawn about the actual efficacy of the intervention. It is recognised that the use of study designs that warrant high internal validity is not always feasible, and that research with lesser degrees of validity can provide clues for further research into interventions for executive disorders. Moreover, our review study showed that in many of the intervention studies insufficient care was taken to determine whether interventions were effective outside the setting in which they were trained, especially in daily life, or after treatment had ended. In other words, the external validity of many studies was unsatisfactory. Nevertheless, the review also showed that a number of intervention studies used adequate research designs to rule out alternative explanations for their positive therapy results. Moreover, some of these studies investigated treatment effects beyond the actual intervention, so that generalisation was assessed either in time or place. Several interventions were examined more than once, providing more stable confirmation of effects.

The experimental treatment described in chapters 5 and 6 of this thesis was intended to introduce a new treatment for executive dysfunction in the field of cognitive rehabilitation. Eligibility for this treatment was not based on lesion location, but on executive problems in daily life, as reported by brain-injured patients and/or their proxies. Also, these complaints were validated by subsequent neuropsychological assessment. Thus, the starting point for participation in this treatment was the clinical diagnosis of executive dysfunction. As a result of the large variety of symptoms that such a broad inclusion criterion entails, this new treatment was designed to embrace the most common executive symptoms. The patients in this multifaceted treatment were trained to cope with everyday executive problems in several executive domains, like planning, initiation and self-monitoring. Such a comprehensive treatment had never been implemented and evaluated before. Our study has shown that it is possible to effectively treat a variety of executive symptoms within one treatment. In addition, the multifaceted treatment was embedded in theoretical frameworks of executive functioning (see chapter 1). Deriving a treatment from these frameworks has several advantages (Burgess & Alderman, 1990; Burgess & Simons, 2005). First, theoretical frameworks provide a clear rationale for the coupling of cognitive and behavioural disorders with therapeutic interventions. Second, theoretical frameworks characterized in terms of cognitive-neuropsychological theory
allow the prediction of implications for therapy. Finally, interpreting cognitive-behavioral problems in terms of information-processing deficits also helps the rehabilitation professional to understand how the process of change might occur in therapy. Moreover, the efficacy study was conducted using a randomized controlled design, so that the internal validity was high. The primary outcome measures were administered by independent assessors who were blind for treatment condition and subjects were blindly and randomly assigned to either the experimental or the control condition. The review study presented in chapter 4 shows that double blind randomized designs are rarely used. Thus, the positive effects of our intervention study described in chapter 5 can reliably be attributed to the experimental, multifaceted treatment. Generalisation of treatment effects was pursued on two levels. First, treatment exercises were designed to optimize behavior in the real world by inviting patients to submit and practice executive problems they encountered in their daily lives. Second, the assessment of treatment effects was targeted at multiple aspects of executive functioning in daily life. Traditional paper-and-pencil neuropsychological tests tapping single executive aspects are unlikely to uncover these effects. To determine the effectiveness of an intervention outcome measures should reflect the intended effects of the intervention, i.e. should be congruent with these effects (Cicerone, 2005). Thus, if an intervention aims to improve everyday executive functioning, the beneficial effects should not only be measured with laboratory-based executive tests, but with measurements at the level of executive activities and social participation. The measurements in our effect study were indications of daily life executive functioning and were thus congruent with the intended effect of the intervention.

Although the outcomes of our intervention study are positive, there is a downside to the treatment that we used. As it is a comprehensive treatment, it is also lengthy and laborious and may therefore be costly. We acknowledge that it may be difficult to implement such a treatment into current clinical practice. A shorter version of the treatment that is less comprehensive might on the one hand still contain the necessary ingredients for patients to benefit in daily life, but on the other hand be more manageable. Such a shorter version of the treatment should at least encompass strengths and weaknesses analysis to promote awareness of deficits, as well as the element of Goal Management Training (Levine et al., 2000; Robertson, 1996), including extra consideration for script generation. On the other hand, although a comprehensive treatment might be lengthy and costly, its multifaceted and comprehensive character might also prove to be efficient, in the sense that treating several executive aspects simultaneously may avoid further care consumption afterward, because some essential executive problems have been neglected.

The study presented in chapter 6 provides supplementary findings to the study described in chapter 5 that established the efficacy of a multifaceted intervention for executive dysfunction after brain injury. The aim was to examine predictors of treatment success. It was investigated whether individual differences in the degree of improvement of the experimental group could be related to differences in scores on three types of predictor variables. These predictor variables were patient characteristics (age and IQ), executive tests (BADS and TOL), and executive questionnaires (DEX and PCRS). The results showed that correlations between predictor variables at baseline and three outcome measures were largely non-significant. Only IQ and TOL correlated with two of the primary outcome measures. The general conclusion is that, with the exception of IQ and TOL, the multifaceted treatment is a highly versatile treatment, suitable for a broad range of patients with executive impairments. The IQ-variable predicted that patients with a lower IQ made significantly more progress in role resumption compared to patients with a higher IQ. However, at baseline patients with a higher IQ had already resumed their social roles to a higher degree than patients with a lower IQ, so that for patients with higher IQ there was less room for improvement. Of the executive test variables, only the TOL showed that patients with better planning abilities benefited more from the multifaceted treatment than patients with lower TOL-planning skills. In short, the treatment described in chapter 5 appears to be beneficial for a large group of patients with executive dysfunction after brain injury.

In this study a total effect size of the multifaceted treatment was calculated at follow-up. This effect-size was large, indicating that treatment effects were still present and substantial six months after treatment. A suggestion for further research concerning the prediction of improvement after treatment is that change is perhaps better assessed using a Reliable Change Index (RCI). A RCI can be calculated taking into account factors like regression to the mean and practice effects, factors that were not accounted for in our study using raw difference scores of three outcome measures (see for example Gehring, Aaronson, Cundy, Taphoorn, & Sitskoorn, 2011). The RCI can subsequently be correlated with the predictor variables.

In summary: only a few studies have investigated both the assessment and treatment of executive disorders as reported by patients and/or proxies. The studies described in this thesis show that it is possible to objectify these complaints using neuropsychological tests, questionnaires, and structured
observations. Script generation tasks also substantiate the everyday executive problems that these patients have. The efficacy study performed with a new, multifaceted treatment aimed at reducing these everyday complaints shows that it is feasible to administer such a demanding intervention to brain-injured subjects. Not only did 75 brain-injured subjects complete the treatment, they also succeeded in bringing about a significant and durable change in their executive behaviour in daily life. This last result should be the primary goal of every neuropsychological intervention: improving independent daily life functioning that is compromised by the devastating consequences of an injury to the brain.

References


Samenvatting

Het onderwerp van dit proefschrift is de diagnostiek en de behandeling van executieve stoornissen bij patiënten met niet-aangeboren hersenletsel. Stoornissen in de executieve functies kunnen ontstaan na een CVA (beroerte), traumatisch hersenletsel, een hersentumor, een hersenontsteking, of na een zuurstoftekort in de hersenen. Onder executieve functies wordt verstaan hersenprocessen van hogere orde die meer basale cognitieve functies aansturen. Executieve functies zijn bijvoorbeeld verantwoordelijk voor doelgericht en strategisch handelen, voor redeneer- en aanpassingsvermogen. Goed werkende executieve functies zijn uiterst belangrijk voor het adequate functioneren in het dagelijks leven: ze maken het een individu mogelijk om zelfstandig en sociaal aangepast te leven. Patiënten die als gevolg van een hersenletsel lijden aan executieve dysfuncties, hebben problemen met het plannen, de initiatiefname, en de monitoring (controlleren) van het eigen gedrag. Deze stoornissen gaan vaak samen met aandachtsstoornissen (bijvoorbeeld verhoogde afleidbaarheid), geheugenstoornissen (met name in het werkgeheugen) en problemen met ziekte-inzicht. Er wordt gesproken van een dysexecutief syndroom wanneer de interne regulatie van gedrag zodanig verstoord is dat iemand afhankelijk wordt van externe structuur; hij/zij kan zijn/haar gedrag niet meer effectief plannen en het tijdig starten en stoppen van gedrag is niet meer mogelijk.

Alexander Luria, één van de grondleggers van de moderne neuropsychologie, was één van de eersten die executieve functies beschreef en ze koppelde aan een specifieke neuroanatomisch gebied, namelijk de prefrontale cortex. Volgens Luria waren deze hersengebieden gesuperponeerd op alle andere corticale gebieden en als zodanig hadden de prefrontale gebieden de functie van programmeren, reguleren en verifiëren van gedrag. Tegenwoordig gaat men niet zozeer uit van afgebakende hersengebieden waar executieve functies gelokaliseerd zouden zijn, maar wordt als neuroanatomisch substraat van de executieve functies een uitgebreid neuronaal netwerk verondersteld, waarin de prefrontale hersengebieden verbonden zijn met andere gebieden, zowel corticale gebieden (bijvoorbeeld de anterieure cingulate cortex en parietale gebieden), als subcorticale gebieden (zoals de basale ganglia), alsmede het cerebellum. In dit proefschrift zijn diverse studies beschreven, die uitgevoerd werden bij een groep patiënten met executieve dysfuncties tengevolge van niet-aangeboren hersenletsel en die verwezen waren voor late poliklinische revalidatie. Late poliklinische revalidatie veronderstelt dat bij patiënten geen spontaan herstel meer valt te verwachten en dat de hersenbeschadiging en de gevolgen daarvan chronisch
zijn. Behandeling in die fase is meestal niet gericht op herstel van functie, maar op compensatie: het omzeilen van een stoornis door gebruik te maken van cognitieve strategieën en/of hulpmiddelen, zodat de patiënt er minder last van ondervindt. Het feit dat de patiënten in huidig onderzoek waren aangemeld voor deze vorm van revalidatie geeft aan dat zij nog voldoende leer- en communicatiemogelijkheden hadden en dat hun executieve en cognitieve stoornissen niet zodanig waren, dat zij geen baat meer zouden hebben bij deze vorm van behandeling. Bovendien werden de patiënten die deelnamen aan de onderzoeken in dit proefschrift niet geklasseerd op basis van frontaal letsel of enkel neuropsychologische tests, maar vond selectie en inclusie plaats op basis van ernstige executieve klachten in het dagelijkse leven, gestructureerde observaties en neuropsychologische tests. Dus, de participerende patiënten hadden executieve klachten (en/of mensen in hun directe omgeving observeerden deze problemen bij hen) tengevolge van hersenbeschadigingen van verschillende aetiologie (CVA, traumatisch hersenletsel, hersentumor, hersenontsteking, of zuurstoftekort), maar niet perse frontaal letsel. Het doel van de studies was om executieve problemen bij een ‘doorsnee’ poliklinische revalidatiepatiënten met executieve klachten te onderzoeken in dit proefschrift niet geïncludeerd op basis van frontaal letsel. Bovendien werden de patiënten die deelnamen aan de behandeling omzeilen van een stoornis door gebruik te maken van een groep gezonde controleproefpersonen. Uit de resultaten van het onderzoek bleek dat de tests, vragenlijsten en observatielijsten sensitief zijn voor de executieve problemen van de patiënten. Er werden geen significante verschillen gevonden tussen de scores op een vragenlijst die executief functioneren meet (de DEX-vragenlijst) van patiënten en hun naasten. Deze bevinding werd geïnterpreteerd als een aanwijzing voor een relatief intact ziekte-inzicht bij de patiënten. Bovendien bleek een combinatie van enkele ‘open-einde’ neuropsychologische tests en de DEX-vragenlijst optimaal voorspellend te zijn voor het onderscheid tussen patiënten en gezonde controles. Dit heeft relevante gevolgen voor het klinisch handelen: middels deze korte batterij van ‘open-einde’ tests en de DEX is kennelijk een snelle en kostenbesparende identificatie van patiënten met executieve stoornissen mogelijk. In Hoofdstuk 3 werd onderzocht in welke mate patiënten met executieve stoornissen problemen hebben met scriptgeneratie. Een script is een in het geheugen opgeslagen complexe ‘kennis-eenheid’ of representatie van een reeks handelingen die nodig zijn om een bepaald doel te bereiken. Het plannen van complex, doelgericht gedrag is afhankelijk van het vermogen dergelijke scripts te activeren. Het verwerken van scriptinformatie wordt verondersteld een van de functies van frontale hersengebieden te zijn. Uit verschillende onderzoeken is naar voren gekomen dat frontaal letsel kan leiden tot stoornissen in het genereren van scripts, die bijvoorbeeld tot uiting komen in het maken van volgordefouten, irrelevante intrusies, en het beëindigen van het script alvorens het doel is behaald. Het huidige onderzoek richtte zich op de vraag of patiënten met executieve stoornissen, vastgesteld volgens een functionele definitie in plaats van op basis van alleen frontale beschadigingen, problemen hebben met het genereren van scripts. De prestaties van de 48 participerende patiënten werden vergeleken met die van een groep van 99 gezonde proefpersonen. De resultaten van het onderzoek lieten zien dat patiënten en gezonden evenveel relevante acties produceerden, maar dat de proportie irrelevante acties bij de patiënten significant hoger was, hetgeen erop wijst dat de patiënten moeite hebben met de inhibitie van irrelevante gedachten. Bovendien noemden de patiënten in de opbouw van hun
Hoofdstuk 4 is een kritische review van de literatuur met betrekking tot de behandeling van executieve stoornissen. Er worden 46 artikels, die 54 interventiestudies omvatten, gecategoriseerd naar type behandeling en onderzocht op interne en externe validiteit. Interne validiteit is de mate waarin de effecten van een onderzochte behandeling daadwerkelijk zijn toe te schrijven aan die behandeling en niet aan onderzoeksartefacten. Wat betreft dit type validiteit zijn de interventiestudies hiërarchisch geordend in functie van hun experimentele design. Methodologisch goed opgezette studies met een design dat meer experimentele controle bevat, hebben een hogere interne validiteit, zodat de effecten met grotere zekerheid aan de behandeling kunnen worden toegeschreven. Externe validiteit heeft betrekking op de mate waarin de behandeleffecten behouden blijven na verloop van tijd en de mate van generalisatie naar niet-getrainde situaties in het dagelijkse leven. Van de 54 interventiestudies was slechts een kwart een ‘true experiment’ met de hoogste interne validiteit, dus een groepsstudie waarbij een experimentele conditie vergeleken wordt met een controleconditie en waarbij de verdeling van subjecten over de condities gerandomiseerd is. Optimaal was ook dat ruim de helft van alle interventiestudies single-case designs hanteerden, ofwel studies van een individu. In meer dan de helft daarvan was de experimentele controle zodanig dat valide conclusies getrokken konden worden met betrekking tot de behandeleffecten. Wat betreft de externe validiteit werd in ongeveer de helft van alle interventiestudies een follow-up meting gedaan, waarbij het verworven behandeleffect altijd behouden bleef. Overdracht van effecten naar het dagelijkse leven werd in meer dan de helft van de studies genoemd, maar slechts in enkele gevallen daadwerkelijk en adequaat gemeten. Samenvattend bleken de effecten van de verschillende interventies veelbelovend; er werd in alle gevallen een positief effect van behandeling gevonden. Echter, de manier waarop dit effect verkregen en gemeten werd, behoort verbetering, zeker als het gaat om de vertaling naar het dagelijkse leven. Uit de huidige review blijkt dat patiënten met executieve stoornissen weldegelijk baat kunnen hebben bij behandeling. Implicaties voor de theorievorming van executieve stoornissen, voor verder onderzoek, en de klinische praktijk worden besproken.

Hoofdstuk 5 beschrijft de evaluatie van een nieuwe neuropsychologische revalidatiebehandeling voor patiënten met executieve stoornissen nagevolgd van hersenletsels. De behandeling was erop gericht om meerdere aspecten van het executieve functioneren tegelijk aan te pakken; de behandeling was dus ‘multifaceted’. Daarnaast werd het belangrijk geacht dat de behandeling toe te snijden was op de specifieke beperkingen van de patiënt. Van essentieel belang was de implementatie van het geleerde in het dagelijkse leven. Zowel de behandeling zelf voorziet hierin, als ook de evaluatiestudie naar de behandeling. Om het effect van de behandeling op het dagelijkse leven te meten, werden naast traditionele neuropsychologische tests ook vragenlijsten, gestructureerde observaties, en een generalisatietaak als uitkomstmaten gekozen. De experimentele behandeling werd vergeleken met een controlebehandeling bestaand uit een gecomponeerde cognitieve functietraining. De patiënten werden gerandomiseerd toegewezen aan één van beide behandelcondities. Uit de resultaten bleek dat beide groepen even tevreden waren met hun behandeling, maar dat de patiënten in de experimentele conditie significant beter functioneerden op executief gebied en dat zij hun rollen in het dagelijkse leven significant beter hadden opgepakt dan de patiënten in de controleconditie. Uit de follow-up meting lieten maanden na afronding van de behandeling bleek dat de effecten behouden bleven.

In Hoofdstuk 6 worden de resultaten gerapporteerd van aanvullende analyses op de behandelbeoordeling uit hoofdstuk 5. Ten eerste wordt een effectsize berekend voor het totale behandeleffect. Deze effectsize is aanzienlijk. Ten tweede wordt onderzocht welke predictoren het behandelsucces in de experimentele groep voorspellen. Er zijn correlaties berekend om vast te stellen of individuele verschillen in de mate van vooruitgang na de behandeling gerelateerd zijn aan verschillen in scores op drie typen predictorvariabelen. De correlaties tussen de baseline predictorvariabelen en de
uitkomstmaten bleken grotendeels non-significant te zijn. Geconcludeerd werd dat de multifaceted behandeling voor executieve dysfunctie van toepassing is op een grote variëteit van patiënten met executieve stoornissen, ongeacht etiologie, leeftijd, intelligentieniveau, of ernst van de executieve stoornissen gemeten met een neuropsychologische test, de Behavioural Assessment of the Dysexecutive Syndrome (BADS). Alleen scores op een planningstest, de Tower Of London test (TOL), bleken in lichte mate voorspellend te zijn voor behandelssluit en daarmee is de TOL wellicht een instrument dat kan helpen bij de selectie en inclusie van patiënten voor de behandeling.

Hoofdstuk 7 biedt een overzicht van de belangrijkste bevindingen van dit proefschrift. Er worden theoretische en methodologische overwegingen besproken en suggesties gedaan voor verder onderzoek en voor de klinische praktijk, zowel wat betreft de diagnostiek als de behandeling van executieve stoornissen bij patiënten met niet-aangeboren hersenletsel. In de conclusie wordt gepleit voor meer ecologisch valide meetmethoden om de ernst van executieve stoornissen te meten en de gevolgen voor het dagelijkse leven te schatten. Wat betreft de behandeling van executieve stoornissen wordt geconcludeerd dat behandeling weldegelijk zin heeft, maar dat er behoefte is aan meer eenduidigheid in type behandeling en een betere verificatie van de onderzochte behandeleffecten middels sterkere onderzoekdesigns. Ook wordt gepleit voor meer aandacht binnen de behandeling voor de implementatie van het geleerde naar de dagelijkse praktijk van de patiënt. Met de nieuwe neuropsychologische revalidatiebehandeling, beschreven in hoofdstuk 5 en 6, werd getracht om meer tegemoet te komen aan deze wensen. De multifaceted behandeling laat zien dat het mogelijk is om meerdere executieve symptomen tegelijkertijd te behandelen, en dat het zinvol is om de behandeling aan te passen aan het niveau van de patiënt en aan de dagelijkse praktijk van de patiënt. De behandeling is bovendien gestoeld op neurocognitieve theoretische modellen en gebaseerd op andere bewezen effectieve behandelingen. Voor de evaluatie van de behandeling werd gebruik gemaakt van een experimenteel design met de hoogste interne validiteit: een Randomized Controlled Trial (RCT). Voor de effectmeting werd gebruik gemaakt van ecologisch valide taken, vragenlijsten, en observaties. De behandel evaluatie heeft daarmee bijgedragen aan het voornaamste doel van de neuropsychologische revalidatie: het optimaliseren van het zelfstandig functioneren in het dagelijkse leven van patiënten met executieve problemen.

Dankwoord

Het is zover, het proefschrift is af. Het was een lang traject en het waren boeiende, leerzame jaren. Ik wil graag van de gelegenheid gebruik maken om op deze laatste bladzijden van het proefschrift die mensen te bedanken die een bijdrage hebben geleverd aan de totstandkoming van het onderzoek en het proefschrift en degenen die mij door dik en dun gesteund hebben.

Luciano Fasotti, jij wist me te overtuigen onderzoek te gaan doen, terwijl ik dat helemaal niet mijn plan was. Ik wilde toch klinisch gaan werken! Het mooie is dat ik nu zowel met het klinische werk, als met onderzoek ervaring heb opgedaan. Wat heb ik veel geleerd! Niet alleen over het verzamelen van data en het schrijven van artikelen en hoofdstukken, het geven van lessen en presentaties, maar ook over de behandeling van patiënten met zulke ernstige aandoeningen. Ik ben je zeer dankbaar dat je altijd vertrouwen in mij hebt gehad en me alle ruimte en tijd hebt gegeven me te ontwikkelen. Ik bewonder je enthousiasme, je passie voor het vak, je volharding en je deskundigheid.

Joke Spikman, ook jij was een fijne begeleider en een bron van inspiratie. Ik ben je dankbaar voor je feedback en je suggesties. Ik vond het geweldig om me te verdiepen in executieve functies en jou als kennisbron te mogen raadplegen. Ik heb veel geleerd over het opzetten en uitvoeren van een multicenter onderzoek, het ontwerpen van een behandelprotocol en het schrijven van artikelen. En alsof dat nog niet genoeg is, schreven we samen ook nog twee boekhoofdstukken. Ik hoop dat we onze samenwerking in de toekomst kunnen voortzetten.

Wiebo Brouwer, ik vond het een hele eer dat je mijn promotor was. Het was heel stimulerend om met behulp van jouw kennis en ervaring van de klinische neuropsychologie dit onderzoek te mogen uitvoeren. Je had altijd verfrissende ideeën tijdens het overleg dat we hadden met betrekking tot het onderzoek.

Roy Kessels, jij werd op de valreep toegevoegd aan het rijtje promotoren. Ik bedank jou voor je stimulerende houding en je optimisme, waardoor zelfs de moeilijkste klussen iets vanzelfsprekends krijgen. Ik ben blij dat je mij overhaalde om op de KNP-opleiding te solliciteren; alweer een stap verder en alweer een gelegenheid om bij te leren!

Ik dank de leden van de manuscriptcommissie voor het nauwgezet lezen van het manuscript: Jos Egger, Caroline van Heugten, en Sander Geurts.
Philippe Allain, merci d’avoir entrepris un long voyage pour être présent à la discussion de ma thèse. Notre collaboration à la rédaction de l’article sur les scripts m’a été très agréable.

Toni Rietveld, wat is een promovendus zonder statistiekhulp? Stuurloos! Ik ben je dankbaar voor je geduldige, heldere manier van uitleggen en je hulp bij het zoeken naar samenhang in mijn data. Bovendien heb ik me goed vermaakt met je geweldige gevoel voor humor en heb ik een aantal prachtige Nederlandse gezegden bijgeleerd (dat is een opgelegd pandoor).

Veel dank gaat uit naar de patiënten en hun families die hebben deelgenomen aan dit onderzoek. Met volle inzet ondergingen zij een intensieve training en lange testsessies. Jullie bijdrage aan de wetenschap heeft ertoe geleid dat er nu een behandelprotocol beschikbaar is dat bewezen effectief is en wordt toegepast in klinieken in heel Nederland.

Ik ben dank verschuldigd aan alle collega-aio’s, projectstudenten en projectmedewerkers die op de een of andere wijze een bijdrage hebben geleverd aan de dataverzameling, de invoering en verwerking van data, en die met hun goede ideeën en kritische blik mij van nieuwe inzichten voorzagen. In het bijzonder wil ik Aglaia Zedlitz, Anneke Hol, Hennerieke Rietberg, Klartje Molleman, en Carly Betting bedanken. Daarnaast natuurlijk Kirsten Lamberts en Marlies van Kessel: ik vond het heel fijn om tegelijk met jullie onderzoek te doen en het lief en leed van een promovendus te kunnen delen. Ook van de bijeenkomsten van het Consortium Cognitieve Revalidatiedielat heb ik veel geleerd en ik wil van de gelegenheid gebruik maken om de leden van het Consortium te bedanken. Chantal Geusgans en Ieke Winkens, wat leuk dat we nog steeds collega’s zijn!

Doordat ik verschillende werkplekken heb (gehad), zijn er ook veel collegae te bedanken, die direct of indirect betrokken zijn geweest bij het onderzoek en het proefschrift. Allereerst bedank de collegae van de Sint Maartenskliniek, alwaar (mijn gedeelde van) het onderzoek plaatsvond. Daarnaast de collegae van de Radboud Universiteit, van de leerstoelgroep Neuro- en Revalidatie-psychologie binnen het DCC. En sinds anderhalf jaar is daar het UMCN St. Radboud als werkplek bijgekomen, alwaar ik weer nieuwe (en oude) collegae ontmoette en weer veel nieuwe dingen leer over ons mooie vak neuropsychologie. Ik ben blij dat jullie me aldaar geholpen en gesteund hebben, door patiënten te verwijzen voor deelname aan het onderzoek en/of door motivatie en betrokkenheid te tonen bij de grote klus die het promoteren is.


Curriculum Vitae

Donders Graduate School
for Cognitive Neuroscience Series


