Executive functioning and ecological validity in fMRI, neuropsychological assessment and rehabilitation

Lamberts, Kirsten Froukje

Chapter 3

A Real-Life, Ecologically Valid Test of Executive Functioning: the Executive Secretarial Task.

Introduction

Information from neuropsychological assessment can serve several purposes including ruling out or confirming diagnoses or deciding which treatment a patient should be offered. Another major purpose is the prediction of a brain-injured patient's functioning in daily life situations: whether he or she will be safe at home, be able to resume previous work, or to resume any other daily life activities. Central to this last purpose is the process of translating test results into expected performance in real-life situations. However, most neuropsychological tests do not provide the information needed to make evidenced-based recommendations, and this issue is probably even more relevant where executive functions (EF) are concerned.

Whilst there have been a number of definitions of ecological validity, in a neuropsychological context it has been defined by Sbordone (1996) as: "the functional and predictive relationship between the patient's performance on a set of neuropsychological tests and the patient's behaviour in a variety of real-world settings" (p. 16). One approach to the development of ecologically valid tests is to use test items or tasks that resemble situations we encounter in everyday life (Burgess et al., 2006b; Chaytor et al., 2003; Spikman et al., 2000a). Of course, simply resembling everyday situations does not guarantee that performance in the test environment will predict behaviour: in real-life this has to be empirically established.

In relation to the estimation of a person's executive functioning in daily life situations, neuropsychological tests have fallen short. Well known tests for EF are: the Trailmaking test, Stroop test (Stroop, 1935), Wisconsin Card Sorting test (WCST; Berg, 1948), Tower of London test (ToL; Shallice, 1982), and the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1996). In this list of tests, the BADS is the most ecologically valid measure widely available. It has been shown to be more sensitive to everyday executive dysfunction than other measures (Bennett, Ong, & Ponsford, 2005a; Wilson, 1993; Wilson, Evans, Emslie, Alderman, & Burgess, 1998), but it is also limited in its ability to predict everyday functioning in patients with brain injury (Norris & Tate, 2000b; Wood et al., 2006). Manchester (2004) suggested that office-based tests of executive functioning might be of use when deciding on rehabilitation approaches, but the assessment of executive deficits in daily life may be best achieved by naturalistic assessment measures combined with information of significant others.

The reason why most neuropsychological tests are not very good at predicting everyday functioning may lie in the fact that many of the tests in use today were developed within a diagnostic tradition (Chaytor et al., 2003), where the primary purpose is to determine whether or not a cognitive impairment is present. Burgess et al. (2006b) discuss how, particularly in relation to executive functioning, this approach has led to the development of tests that were not good at predicting actual behaviour in everyday life.

Executive functioning is broadly defined in terms of problem-solving, a concept which incorporates processes such as planning, goal formation, and goal management (Duncan, 1986; Levine et al., 2000; Shallice & Burgess, 1996). Everyday executive functioning involves maintenance of multiple goals and sub-goals, with priorities that change over time requiring self-initiative, self-monitoring, and self-regulation. A test which aims to predict everyday executive functioning must therefore be sensitive to these types of cognitive ability.

The Executive Secretarial Task (Spikman et al., 2009) was designed using a function-led approach as recommended by Burgess et al. (2006b). These authors state that the development of most tests has been based on a theoretical construct or cognitive resource. Those tests provide information on hypothesised resources which is not easily translated into real-life functioning. Especially where such complex functions as EF are concerned, test development based on a function, or the directly observable behaviour is most likely to result in a test fit for clinical purposes. A function-led approach in test development would for instance ensure that the actual situation in which the assessment takes place is more like daily life and that more elements involved in the complex process of EF are being assessed. The EST extends over a longer period of time than most traditional EF tests, requiring organisation and prioritisation of multiple tasks, while delayed intentions, interruptions and deadlines have to be dealt with. In this first phase of studying the EST, we examined whether it discriminated between healthy controls and brain-injured patients known to have deficits in EF. In subsequent studies we will examine whether the EST also discriminates between brain-injured patients with and without EF problems. Furthermore, we expect the EST to be able to clarify which components of the EF are most damaged. With that information neuropsychologists should be able to predict performance in a specific daily life situation. This in turn should lead to useful recommendations to overcome problems or to suggestions on the main issue within a treatment.

Methods

Participants

The task was administered to 35 brain injured patients and 57 healthy controls (table 1). Participants in the healthy control group were recruited via advertisements in local newspapers. The patients were participants in a study evaluating a new multifaceted treatment of executive dysfunction (Spikman et al., 2009). The EST was part of the six-month follow-up measurement in that study. In Spikman et al.'s (2009) study, the patients were randomly assigned to either the experimental treatment group or the control treatment group. The reason to include a control intervention instead of a no-treatment group was the avoidance of differences in expectations between both patient groups. In the present study the experimental group was excluded as those patients had been exposed to an intervention specifically designed to improve executive functioning. The control group had participated in a computerized training (CogPack) aimed at improving cognitive functioning in general. This training was not expected to result in improvements of executive functioning any more than resuming daily life without treatment would. Data supported this assumption (Spikman et al., 2009). The study took place in seven rehabilitation centres and two academic centres in the Netherlands. In order to be included the patients had to have Acquired Brain Injury (ABI) of a non-progressive nature, for instance TBI, stroke or cerebral tumours, with a minimal time post onset of three months and no maximum. Furthermore, they had to be in the age range of 17-70 and had to live in their home situation, that is, they should not be hospitalized or living in a sheltered environment. Finally, patients had to be referred for outpatient rehabilitation treatment because of post-injury problems of a clearly dysexecutive nature, either reported by themselves or observed by others. These problems had to hamper the resumption of previous activities and roles. If according to the judgment of the rehabilitation physician, using a checklist for dysexecutive problems, patients fulfilled these criteria, verbal and written information about the study was provided to patients and proxies. If patients gave their informed consent to participate, they subsequently underwent a neuropsychological examination to determine whether definitive inclusion could take place.

Measures

The **Dysexecutive Questionnaire** (DEX; Burgess et al., 1996) was used to determine the presence of executive symptoms in everyday life. One version of the questionnaire was completed by the participant, another was completed by a proxy, i.e. someone who knew

the participant well (usually, but not always the partner). A third copy was completed by a therapist (usually a neuropsychologist) who was not further involved in this study.

The **Executive Observation Scale** (EOS) is a translation of the observation list developed by Pollens (1988). It consists of eight 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 scale from 1 (representing complete inability) to 4 (representing completely independent and able) and the total score therefore ranges from 8 to 32. The EOS was completed by a therapist, who had knowledge of the patient's performance in daily life situations and was not involved in the EST assessment.

In order to measure executive functioning on a participation level, the **Role Resumption List** (Brand & Brouwer, 2002; RRL; Spikman, Brand, & Brouwer, 2003) was administered. This is a structured interview making an inventory of objective changes, compared to premorbid level, in quality of activities and time spent on four areas of everyday life (vocational functioning, leisure activities, social interaction with partner and family, and mobility). This interview was administered by a neuropsychologist not further involved in the study. Through observations and the information gathered during the interview this therapist got to know the patients well enough to additionally complete the EOS and the DEX. Based on the interview, each of these four areas is rated on a five-point scale, judging whether and how much the patient's role-behaviour has changed since brain injury. It includes a return to work scale similar to the one used by Van Zomeren and Van den Burg (1985), whereby a score of 0 indicates that former work or study have been resumed without changes, and a score of 4 indicates "not working at all". For the other subscales a score of 0 represents no change and a score of 4 represents a severe loss of independence compared to premorbid functioning. The scores for the four domains were added up to a total score.

Two non-executive tests were included to provide information on cognitive functioning. The time score on the **Trail Making Test** part A (TMT A) was used as an indication of information processing speed. Memory was tested with the **15 Words test**, a Dutch version of Rey's Auditory Verbal Learning Task (Deelman, Brouwer, Zomeren van, & Saan, 1980). The 15 Words test total score was used in analysis.

The **Behavioural Assessment of the Dysexecutive Syndrome** (BADS; Wilson et al., 1996) was administered in its Dutch translation (Krabbendam & Kalff, 1997) as it is the best available test battery for EF. This test was designed as an ecologically valid test of EF and consists

of six subtests, all tapping different aspects of EF. These are: the Rule Shift Card Test (RS), the Action Program Test (AP), the Key Search Test (KS), the Temporal Judgement Test (TJ), the Zoo Map Test (ZM), and the Modified Six Elements Test (MSET). A parallel version for the Zoo Map was used in the Spikman et al. study (2009) to prevent re-test effects: the Shopping Centre test (SC; Spikman & Landskroon, 2000b). Therefore some of the participants in the present study had also been assessed on the SC. The designs had the same difficulty level which was confirmed when comparing the scores on the ZM and SC in a large group of healthy elders (155 participants; mean age 74; data collected within the Groningen Intervention Study on Successful Aging). We used the summary BADS age score in our analyses.

Finally, the **EST** (Spikman et al., 2009) requires organization and prioritisation of multiple tasks over a long time span, while dealing with delayed intentions, interruptions and deadlines. In this three hour-task a job assessment procedure is simulated. It is comparable to the Multiple Errands Test (Alderman et al., 2003; Shallice et al., 1991) or the Hotel Task (Manly et al., 2002). Participants are alone in a room with a box containing several simple secretarial assignments (see Figure 1). Furthermore, they are provided with a phonebook, calculator, telephone, some office supplies, a list with company rules, a planning aid in the form of a day agenda, and a map showing the participant's office and neighbouring offices. The assignments have to be organized, initiated and executed; some of them with a deadline. Examples are: filling in zip codes on envelopes and posting them in time for the external post round; counting the supplies and replenishing the stocks by delivering the order form in time at the right place; or searching for suitable restaurants for the company dinner in the phonebook. A unique feature of this test is that, unlike most other EF tasks, it taps self-initiation: the instruction does not provide more information than that the participant can find assignments in the box and that these all have to be carried out. No further cues are provided on how or when to carry out these assignments. Necessary materials and information are all available, but have to be actively searched for. For example, in one of the assignments the participant has to make a travel schedule for specific dates and destinations. This can be done by using the telephone (and for which the instruction for use can be found in the list with company rules), but the participant is not explicitly told to use the phone for this purpose. At fixed times only, questions can be asked to the "manager". During the assessment, the participant is interrupted with an urgent new assignment. The task yields three scores (for details on calculation see Figures 2 and 3): Initiative, reflecting all the actions the participant has initiated without being told so; Prospective, reflecting all the actions that were correctly carried out on a later moment; and Executive,

reflecting all the actions that were correctly carried out at all. Together these scores form the total score.

All measures were part of the large battery of tests used in the study by Spikman et al. (2009). The patients in that study underwent three assessments and the healthy controls only one. In order to avoid the possibility of practice effects, the EST was only administered at the follow-up time point. The EOS, RRL, TMTA and 15 Words test were not presented to the healthy control group. In the present study, we used the patients' TMTA, 15 Words test, BADS and DEX baseline scores, and the EST, EOS and RRL scores from the third, follow-up, measurement.

Statistical analyses

The two groups of participants were tested to determine whether they were matched on age and educational level. The normality of test scores was examined to determine whether it was appropriate to use parametric statistical tests. A t-test was used to test whether groups differed from each other on the neuropsychological measures. Finally, in the patient group correlations were determined between the EST scores and the BADS, DEX, EOS, and RRL scores and between the BADS, the EOS and the RRL scores. We excluded cases if they had a missing value.

Results

In total, 92 participants were included: 35 patients, and 57 healthy controls. Table 1 shows the characteristics of the two groups. The groups were well matched on age and educational level. In the distribution of men and women, both groups differed slightly. This had no effect on test results. In the patient group etiology and months since brain injury are also presented. With an average of 45 months since brain injury it is clear that these patients were in the chronic phase after their injury. Most of the patients suffered either a stroke or TBI. Other etiologies included postanoxic encephalopathy or encephalitis.

Table 1 *Demographics of Participants*

	Patients (P) n=35 M *(sd/ range)*	**Healthy controls (HC) n=57** M *(sd/ range)*	**Significance**
Age	44.4 *(15.1 / 17-64)*	47.8 *(11.4 /24-69)*	.082
Edu	4.8 *(1.2 / 2-7)*	5.3 *(1.0 / 3-7)*	.516
M/F %	63 / 37	40 / 60	.02
Months	45.8 *(64.3 / 4 - 288)*	NA	NA
Etiology %	31 / 54 / 14	NA	NA

Edu = educational level (determined using the seven point scale of Verhage (1964) which runs from less than six years primary school (1) to university level (7)); Months = number of months since brain injury; Etiology = TBI / stroke / other

All measures showed highly significant differences between healthy controls and the brain-injured patients (table 2). As expected the BADS and the DEX differentiated very well between patients and healthy controls. Especially the therapists' DEX was a successful measure. This might have been expected as the therapists participating were all experienced and would not have much trouble recognising EF problems. Also the EST proved to differentiate between both groups: not only on the EST total score, but also on the different subscales.

Table 2 *Scores of healthy controls (HC) compared to patients (P)*

	P M *(sd)*	**HC** M *(sd)*	**HCvsP** (1-tailed) T	p	Cohen's d
BADS	85.8 *(14.0)*	102.3 *(12.3)*	5.9	***	1.25
DEX patient	30.8 *(14.0)*	18.3 *(8.6)*	-4.7	***	1.08
DEX proxy	31.6 *(15.7)*	18.1 *(9.9)*	-4.4	***	1.03
DEX therapist	35.3 *(11.7)*	10.1 *(6.6)*	-11.7	***	2.65
EST total (0-45)	28.3 *(10.3)*	37.4 *(5.6)*	4.8	***	1.10
EST initiative (0-13)	9.5 *(3.2)*	11.5 *(1.9)*	3.4	***	0.76
EST prospective (0-8)	5.2 *(2.2)*	6.7 *(1.5)*	3.7	***	0.80
EST executive (0-24)	13.6 *(6.2)*	19.1 *(3.5)*	4.8	***	1.09

***p<.00

In the patient group the EST total score showed a significant correlation with the Behavioural Assessment of the Dysexecutive Syndrome (BADS), the Dysexecutive Questionnaire (DEX) proxy, and the DEX completed by the therapist (table 3). This showed concurrent and ecological validity of the EST in a clinical population. The BADS correlated significantly with DEX-therapist and this correlation was stronger than that between the EST and the DEX-therapist. The BADS did not correlate with the DEX proxy, though the level of correlation was only just below that between the EST and DEX proxy. Both the EST and the BADS showed no significant correlation with the patient's DEX. This is most probably due to a lack of insight of the patients something that has been a consistent finding in previous studies with the BADS (Bennett, Ong, & Ponsford, 2005b; Burgess et al., 1996). To test for discriminant validity, the correlations of the EST with the 15 Words test (.233), and the Trail Making Test part A (-.214) were determined. No significant correlation was found.

Table 3 *Correlations BADS, DEX and EST in the patient group*

	EST total	**BADS**	**DEX patient**	**DEX proxy**	**DEX therapist**
EST total	1	.438(**)	-.022	-.305(*)	-.290(*)
BADS		1	-.252	-.203	-.357(*)
DEX patient			1	.542(**)	.526(**)
DEX proxy				1	.455(**)
DEX therapist					1

** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

We determined correlations between the EST and the Executive Observation Scale (EOS) (table 4) and Role Resumption List (RRL) (table 5) to test whether the EST could indeed serve as an indicator of which component of EF is most damaged and whether the EST had predictive value for daily life functioning. The EST correlated significantly with most of the subscales of the EOS and in general higher correlations were found between subscales of the EST and EOS that aim to measure the same component of EF. These results indicate that the EST is able to predict performance in daily life situations and provides a therapist with useful information on which components of EF need most attention in treatment of a patient. In comparison the BADS appeared to correlate less with the EOS (table 6), but the levels of correlations are broadly similar to those between the EST and the questionnaires. In some cases relations between the BADS and EOS were found between subscales designed to measure the

same component of EF, but in other cases these expected correlations were not significant. For instance, the BADS ZM/SC showed no significant correlation with the planning scale of the EOS.

The EST showed significant correlations with several scales of the RRL. In particular, the EST-executive subscale proved to have predictive value as it correlated significantly with three of the areas of everyday life measured by the RRL. This suggests that the EST can give an indication of executive functioning at a participation level. None of the EST subscales showed a significant correlation with the vocational score on the RRL. Also in the case of the RRL the BADS appeared to show lower levels of correlation (table 7), but again the correlations are broadly similar to those found between the EST and the RRL.

Table 4 *Correlations EST and EOS in the patient group*

	EOS awareness	**EOS planning**	**EOS goal setting**	**EOS self-initiation**	**EOS self-inhibition**	**EOS self-monitoring**	**EOS ability to change set**	**EOS strategic behaviour**	**EOS total**
EST initiative	.162	.381(*)	.324(*)	.314(*)	.170	.283(*)	.185	.418(**)	.384(*)
EST prospective	.091	.300(*)	.198	.035	.366(*)	.047	.275	.159	.266
EST executive	.096	.523(**)	.419(**)	.358(*)	.081	.208	.138	.367(*)	.376(*)
EST total	.127	.497(**)	.395(**)	.320(*)	.180	.223	.199	.384(*)	.402(**)

** Correlation remains significant after Bonferroni correction
** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

Table 5 *Correlations EST and RRL in the patient group*

	RRL vocational	**RRL social**	**RRL leisure**	**RRL mobility**	**RRL total**
EST initiative	-.188	-.326(*)	-.160	-.177	-.299(*)
EST prospective	-.176	-.160	-.191	-.157	-.232
EST executive	-.143	-.332(*)	-.303(*)	-.310(*)	-.373(*)
EST total	-.182	-.335(*)	-.273	-.275	-.367(*)

* Correlation is significant at the 0.05 level (1-tailed).
After Bonferroni correction no significant correlations were found.

Table 6 *Correlations BADS and EOS in the patient group*

	EOS awareness	EOS planning	EOS goal setting	EOS self-initiation	EOS self-inhibition	EOS self-monitoring	EOS ability to change set	EOS strategic behaviour	EOS total
BADS RS	-.042	-.107	-.366(*)	-.074	-.360(*)	-.096	.017	-.252	-.218
BADS AP	-.206	.187	-.111	-.146	-.094	.015	.005	-.031	-.060
BADS KS	-.005	.314(*)	.164	.004	.216	.130	.154	.122	.171
BADS TJ	-.121	-.106	.140	.284(*)	-.131	-.144	-.155	.099	-.028
BADS MSET	.122	.131	.122	.258	.206	.131	.277	.355(*)	.291(*)
BADS ZM/SC	-.148	-.119	.083	-.266	.009	.235	-.091	.201	-.032
BADS total	-.066	194	.240	.057	.298(*)	.068	.146	.301(*)	.211

** Correlation is significant at the 0.01 level (1-tailed).
After Bonferroni correction no significant correlations were found.

Table 7*Correlations BADS and RRL in the patient group*

	RRL vocational	RRL social	RRL leisure	RRL mobility	RRL total
BADS RS	.183	.078	.279	.142	.225
BADS AP	.060	-.095	.044	-.087	-.030
BADS KS	-.216	-.229	-.223	-.131	-.276
BADS TJ	.261	-.057	.209	.079	.155
BADS MSET	-.205	-.229	-.293(*)	-.134	-.296(*)
BADS ZM/SC	-.025	.312(*)	.043	.150	.177
BADS total	-.219	-.150	-.241	-.014	-.216

* Correlation is significant at the 0.05 level (1-tailed).
After Bonferroni correction no significant correlations were found.

Finally, sensitivity and specificity of the EST were analysed using various cut-off scores for the EST. The EST resulted in clear differences between both groups and a cut-off score for maximum specificity and sensitivity (table 8) could be suggested: 34-35. With the cut-points examined, optimising sensitivity and specificity still means that a proportion (around 25%) of the healthy controls was below the cut-off. However, within the healthy control population there will be some variation in executive functioning, comparable to IQ, and so 100% specificity would not be expected.

Table 8 *Possible cut-off scores for the EST*

EST total score	**Sensitivity**	**Specificity**	**False Positive**	**False Negative**
22	28.6%	100%	0%	71.4%
26	37.1	94.7	5.3	62.9%
28	45.7	87.7	12.3	54.3
30	57.1	82.4	17.6	42.9
32	68.6	78.9	21.1	31.4
33	69.5	77.2	22.8	30.5
34	70.5	75.4	24.6	29.5
35	71.4	71.9	28.1	28.6
36	80.0	63.2	36.8	20.0
39	87.9	45.6	54.4	12.1

Discussion

The results of this study suggest that the Executive Secretarial Task (EST) gives valuable information on the executive functioning of brain-injured patients in daily life situations. The test proved to discriminate between healthy controls and brain-injured patients with deficits in executive functions (EF) and it gave indications on what aspects of EF were most damaged in the patient group. We can therefore conclude that this first phase of studying the use of the EST in a clinical population suggests that the EST can provide the neuropsychologist with very useful information, in the diagnostic phase as well as with indication for treatment and when deciding on specific goals for treatment. Above all, we demonstrated that scores on the EST are associated with executive functioning in daily life situations. This research has provided enough positive evidence to continue studying this test with the eventual goal of implementing it in clinical practice.

The fact that this was the first time the EST was administered to a clinical population led to several suggestions for improvement of the test and to future research options with the EST. First of all, the EST showed it could discriminate between healthy controls and brain-injured patients with EF problems. The next step in the process of testing its use for clinical practice would be to test whether the EST can also discriminate between brain-injured patients with and without EF problems. This would also provide information on the question whether scores on the EST potentially might have been sensitive to multiple deficits in the patient group and not just dysexecutive functioning. The EST is designed in such a way that this is not likely to be the case: there is no time pressure, the tasks are not intellectually challenging, and memory (with the exception of prospective memory) does not play a role as the instructions can be consulted

whenever the participant wants to. Another positive sign was of course the EST scores showing no significant correlation with scores on the tests used to assess memory and information processing speed. Whilst there was no significant correlation between these measures and the EST, further examination of the specificity of the EST in relation to detection of deficits in executive functioning would be useful. Second, predictive value of the EST in daily life situations should be tested on a more practical level, not only through questionnaires. For instance, is it successful in the process of selecting which job or other daily life activities a brain-injured patient can resume? Results of the present study suggest that on many activities the EST can give valuable information, but the modest correlations with the Role Resumption List (RRL) do not allow firm conclusions on this subject. It seems that especially where vocational functioning is concerned, the EST did not provide sufficient information. There are two plausible explanations for that. To begin with, all unpaid vocational activities like voluntary work or being a housewife are not registered by the RRL. For many patients this was in fact the work they returned to after brain injury. Besides that, for most patients work is the last of the four activities in daily life, measured by the RRL, they start with again. Most of them might not even have started thinking about resuming their work again. It is likely that in a later phase of recovery, when patients have started working again, the EST does show predictive value. To test these hypotheses, information on unpaid work should be added and data on vocational functioning should be collected even longer after ending rehabilitation treatment. Third, in the present study the choice was made to compare baseline scores on the TMTA, 15 Words test, BADS and the DEX with follow-up scores on the EST, RRL and Executive Observation Scale (EOS). The reason behind this was that the EST was expected to give most valuable information at follow-up. To avoid a retest effect, for which especially EF tests are notorious, it could only be administered once. The BADS was needed for inclusion, therefore it had to be administered at baseline. Comparing the EST with the BADS score at follow-up (for patients the third confrontation with the BADS) was no option, again because of the retest effect. One could argue that the time lag had an effect on the results and further study should take this into account. However, as we only included patients who were presented with the control treatment and healthy controls, we did not expect changes in executive functioning over six months. Furthermore, all patients were in the chronic phase, so spontaneous recovery resulting in improvement of test performance over time was not likely. Finally, future validation studies could include a test of prospective memory, such as the CAMPROMPT (Wilson et al., 2005), to analyse its impact on task performance and it would be useful to examine whether past experience of secretarial work has an impact on performance.

Working with the EST also revealed some disadvantages of the test which in most cases are inevitable in the measurement of real-life executive functioning. An obvious disadvantage of the EST is that it is a long test, but this is inherent in a measure of real-life executive functioning. In daily life, plans have to be kept active over time and whether prospective memory is limited or not can only be tested with a test that also stretches over a longer period of time. It would be a challenge though to adapt the EST for use in time-limited clinical practice. Another disadvantage falling in the same category of inevitable disadvantages is that the EST does not seem suitable for retesting. We did not test this here, but as EF are only really addressed in new situations where a person has to generate alternative behaviour instead of the usually performed routine behaviour it seems obvious that a test can only give reliable results the first time it is administered. Therefore, retesting, for instance to keep track of a person's improvements in EF after treatment, can only be done by a specially designed parallel test. This might be an interesting next challenge in the process of developing an ecologically valid test for EF.

A very practical issue that arose while working with the EST was the possibility of using a computer linked to the internet. The absence of this possibility did not give any problems in this population as they were all still able to find the required information in a phonebook. However, some participants explicitly asked whether there was a computer they could use.

The tasks

In box:

- Envelope: internal mail : post these letters on time (room numbers have to be added)
- Envelope: external mail: post these letters on time (postal codes have to be added)
- Envelope: look up train departure times for specified dates and locations (solved by phoning travel information line)
- Envelope: supply stock to the needed amounts. Deliver stock supply form on time at stock suppliers (there is a fixed budget for this. The price list (in folder 'important information') shows the prices. A certain amount remains and patient can chose to spend it on one of two larger purchases (only one can be bought for that amount of money).
- Envelope: find as many restaurants in 5 towns in the province of Friesland where you can enjoy a good meal (no fast-food) with ca. 60 persons. Make a list for the company trip: telephone book or yellow pages can be used. Write down names, addresses and telephone numbers of these restaurants.
- Sheet of paper: Write down what you have done on this agenda.

Interruption 1.5 hours after starting the task (sooner if patient works very fast):

- Urgent: deliver note at (employee two offices down the hall) with message: appointment with Mrs. Kolenbrander at 11.30 is cancelled. This message has to be delivered before 11.30.

Figure 1. An overview of the secretarial assignments a participant is asked to do.

Scoreform EST part I			
Name:	**Number:**	**Date:**	
Task	Action	Score	Notes
1. Internal mail:		**1. Internal mail**	**1.**
A. Room # added (2 pnt)	initiative	**A.**	
B. Posted on time (2 pnt)	prospective	B.	
C. Posted in correct box (1 pnt)	executive	C.	
2. External mail:		**2. External mail**	**2.**
A. Usage phonebook own initiative (2 pnt)	initiative	A.	
B. Posted on time (2 pnt)	prospective	B.	
C. Postal codes added correctly (3 pnt)	executive	C.	
D. Posted in correct box (1 pnt)	executive	D.	
3. Travelling schedule:			
A. Phoned travel information or tried to (3 pnt) *1 *2		**3. Travelling schedule** A.	**3.**
B. Schedules are correct (1 pnt per correct schedule, max. 3 pnt)	initiative executive	B.	
4. Stock:			
A. Usage price list own initiative (2 pnt) ***2**		**4. Stock** A.	**4.**
B. Delivered on time (2 pnt)	initiative	B.	
C. Delivered at stock suppliers(1 pnt)	prospective	C.	
D. Supplied stock correctly (4 pnt) ***3**	executive executive	D.	
5. Company trip:		**5. Company trip**	
A. Usage yellow pages own initiative (2 pnt) ***2**	initiative	A.	
B. > 4 restaurants, in > 2 towns (2 pnt)		B.	**5.**
C. Written down restaurants for larger groups (2 pnt)	executive executive	C.	
6. Interruption:			
A. Answered phone (2 pnt)		*6. Interruption*	
B. Message is delivered on time (2 pnt)	initiative	A.	
C. Message is delivered at correct door (1 pnt)	prospective executive	B. C.	**6.**
7. Agenda: Written down which task is done when (1 pnt per task, max. 6 pnt)	executive	**7. Agenda** A.	**7.**

***1** *Add accurate description how pat has performed this task*
***2** *Asking a question forcing the test leader to give a hint results in 1 pnt reduction, asking for confirmation of own idea no reduction*
***3** *every mistake: 1 pnt reduction*

Figure 2. Score form for all individual assignments.

Scoreform EST part 2 ***Name:*** ***Number:*** ***Date:***

	Add up scores	Max. score	Participant score
initiative	1A, 2A, 3A, 4A, 5A, 6A	13	
prospective	1B, 2B, 4B, 6B	8	
executive	1C, 2CD, 3B, 4CD, 5BC, 6C, 7	24	
total	All tasks	45	

Evaluation test leader

1. What is your impression of the adequateness of the asked questions?

not adequate 1 – 2 – 3 – 4 – 5 very adequate

2. What is your impression of the adequateness of the planning?

not adequate 1 – 2 – 3 – 4 – 5 very adequate

3. What is your impression on how much effort the participant had to put into this test?

a lot 1 – 2 – 3 – 4 – 5 hardly any

4. What is your impression on the overview the participant had?

a lot 1 – 2 – 3 – 4 – 5 hardly any

Questions for participant

1. Own impression on how well the tasks have been performed

not adequate 1 – 2 – 3 – 4 – 5 very adequate

2. Did you have enough time?

far to little 1 – 2 – 3 – 4 – 5 far too much

3. What was your experience of time passing?

passed slowly 1 – 2 – 3 – 4 – 5 passed quickly

4. How much distraction did you experience?

a lot 1 – 2 – 3 – 4 – 5 hardly any

5. How exhausting was the task?

very 1 – 2 – 3 – 4 – 5 hardly

Any other complaints or observations?

Figure 3. Score form for total score and observations of both test leader and participant.

J.M. Spikman D. Boelen, K.F. Lamberts, W.H. Brouwer and L. Fasotti

(submitted for publication)