J Head Trauma Rehabil
Vol. 29, No. 2, pp. 136–146

Assessment of Subcomponents of Executive Functioning in Ecologically Valid Settings: The Goal Processing Scale

Tatjana Novakovic-Agopian, PhD; Anthony J.-W. Chen, MD; Scott Rome, MD; Annemarie Rossi, OTR/L; Gary Abrams, MD; Mark D'Esposito, MD; Gary Turner, PhD; Ryan McKim, PsyD; Jim Muir, PhD; Nancy Hills, PhD; Cathy Kennedy, MS; John Garfinkle, MS, SLP; Michelle Murphy, PsyD; Deborah Binder, MS; Holli Castelli, OTR/L

Objectives: To validate a new functional assessment tool, the Goal Processing Scale (GPS), and to apply it for testing for sources of dysfunction in patients with acquired brain injury. Determining which component processes of executive functioning underlie poor performance in complex, low-structure settings would be valuable for the assessment of deficits and for evaluating the effectiveness of treatments. **Participants:** Nineteen individuals with chronic acquired brain injury (mean age = 41.4 years; chronicity: 6 months to 39 years). **Main Measures:** Two functional assessment tasks: (1) GPS, which evaluates functional performance in the context of achieving a goal in a "real-world" setting, with rating scales measuring overall performance and 8 subdomains of executive functioning; (2) Multiple Errands Test, an unstructured assessment of ability to adhere to rules and complete multiple "real-world" tasks in a short time; and (3) a neuropsychological battery. **Results:** Intraclass correlation coefficients for 2 independent raters ranged from 0.75 to 0.98 for the GPS overall composite score and the subdomain scores. Performance on GPS overall and several subdomain scores correlated with performance on the Multiple Errands Test. Working memory and learning/memory neuropsychological measures predicted functional performance as measured using the GPS. **Discussion:** The GPS shows high interrater reliability, suggesting convergent validity with an established functional performance measure, and produces useful information regarding strengths and weaknesses in different subdomains of executive functioning. Working memory and learning/memory appear to be key determinants of goal-directed functioning for these individuals with brain injury. **Key words:** *brain injury rehabilitation*, *ecologically valid assessment*, *executive functions*, *functional performance*

Some of the most common and persistent sequela of brain injury are deficits in executive control functions, including the abilities of paying attention, holding information in mind, planning, organizing, and developing efficient strategies for completing activities in daily life.[1–9] Executive control deficits can affect many aspects of personal functioning, such as the pursuits of educational and occupational goals as well as numerous aspects of daily life, directly contributing to

Author Affiliations: *Veteran's Administration Medical Center, San Francisco (Drs Novakovic-Agopian, Chen, Abrams, McKim, and Murphy and Mss Rossi and Binder); Veteran's Administration Northern California Health Care System, Martinez (Drs Novakovic-Agopian, Chen, D'Esposito, and Muir and Ms Binder); California Pacific Regional Rehabilitation Center, San Francisco (Drs Novakovic-Agopian and Rome; Mss Rossi, Kennedy, and Castelli; and Mr Garfinkle), University of California, San Francisco (Drs Novakovic-Agopian, Chen, Abrams, and Hills); University of California, Berkeley (Drs Chen and D'Esposito), California; and Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada (Dr Turner).*

This material is based upon work supported by the Office of Research and Development Rehabilitation R&D Service Department of Veterans Affairs, and the California Pacific Medical Center (CPMC) Foundation. The authors thank participating patients and a number of individuals who made this study possible, in particular Gerald Carlin, OTR/L, for help with earlier versions of functional assessment tool and with participant recruitment; Richard Fitzsimons, MS, for conducting study neuropsychological evaluations; Guiomar Scheid, for help with participant scheduling; and Byron Morgenroth, MS, for help with organizing the portions of the study conducted at the CPMC.

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.headtraumahab.com).

Corresponding Author: *Tatjana Novakovic-Agopian, PhD, SFVA Medical Center, 4150 Clement St, Mail Code 127, San Francisco, CA 94121 (tna@cns-site.com).*

DOI: 10.1097/HTR.0b013e3182691b15

poor outcomes.[10–13] Despite their importance, deficits in these "control processes" are often poorly addressed. One roadblock may be that the nature of the functional deficits remains poorly understood, with inadequate tools for quantitative assessment to characterize each individual's functional strengths and weaknesses.

The construct of executive control, referring to abilities important for guiding behavior based on goals, consists of a complex combination of component functions. There are 2 opposing challenges to the assessment of executive functioning. On the one hand, clinical evaluations (including neuropsychological testing) are often inadequate for the detection of dysfunction in these higher-level domains, which are often most apparent in real-life settings. Executive control functions are arguably most engaged in the low structure of real-world settings, where environmental or situational context may not be sufficient to guide behavior. Most conventional neuropsychological tests of executive function are not designed to capture the complexity of responses required in the many multistep tasks that are part of daily life.[14–16] Traditional neuropsychological tests help in identifying specific domains of cognitive dysfunction, but because of their highly structured nature, they may not fully capture or predict an individual's performance in unstructured and complex real-world settings and/or detect deficits that may nevertheless be disabling.

On the other hand, although the importance of more ecologically valid assessment is increasingly recognized,[14–16] measures of component processes that contribute to the efficiency or effectiveness of functional goal-directed behavior in real-world situations are not currently available. Furthermore, assessments in real-world situations are complicated by challenges such as measurement precision and reliability. Improved ability to define each individual's strengths and weaknesses on subcomponents of executive functions in complex real-life environments would be an invaluable advance in the assessment of deficits, planning interventions, and evaluating the effectiveness of treatments.[17]

A number of approaches have been taken for assessing functional performance in ecologically valid fashion. Directly asking patients how they are functioning in the real world has been the most common method for assessing real-world outcomes.[18] Self-report measures have been used to characterize everyday functioning in the areas of memory, communication, and higher-order problem solving.[19] However, they often do not provide a clear relation to real-life performance.[20] A number of studies have demonstrated that self-report is susceptible to biases based on the individual's mood and cognitive status.[18] Another common approach is obtaining input from a knowledgeable informant (such as spouse or caregiver) about an individual's real-life functioning. However, although valuable, this information can be incomplete or biased on the basis of the informant's familiarity and relationship with the patient.[18]

Arguably, the most valid determination of everyday functioning would be direct observation of the person in the real world. Several ecologically valid functional tasks have been developed to assess activities of daily living (ADL). Independent Living Scales[21] was developed for the assessment of instrumental ADL such as simple money management, dialing a phone number, writing a check, and understanding simple instructions. In the Naturalistic Action Test,[22] individuals are asked to accomplish a simple functional task such as making a sandwich. Performance is rated in terms of subtasks accomplished and errors made. These measures, although ecologically valid, are task specific and address simpler ADL.

More complex naturalistic tests such as the Multiple Errands Test (MET),[23] which incorporates multifaceted life-like challenges including multitasking, have been developed in recent years. The MET is an unstructured functional test that assesses a patient's ability to follow outlined rules and complete multiple "real-world" tasks in a limited time. Several versions of the MET have been reported in the literature (including those involving a shopping center or a hospital campus).[23–25] Participants are presented with written instructions and a (shopping center/hospital) map and asked to complete a number of subtasks, while following specified rules. The MET has been shown to capture an ecologically relevant summary of performance on a complex task, as reflected by the number of task rule breaks and task failures. However, the MET is limited in that it lacks specificity with regard to the cognitive subprocesses that contribute to dysfunction.

To help address this gap, we developed an instrument that assesses functional performance on subdomains of executive function necessary to perform a complex "real-life" functional task (the Goal Processing Scale [GPS]). In designing the GPS, we took into account that individuals with executive dysfunction can have difficulties with different cognitive processes necessary for successful goal/task completion–those that are involved in planning a task, actual task execution, or both.[26,27] For example, Duncan[28] described the concept of "goal neglect" where individuals can plan and verbalize their goals and subtasks but fail to maintain goal-directed behavior when executing a task. Goal neglect, or a failure in maintaining the goal and/or goal-relevant information in working memory, can occur as a consequence of being distracted from a goal by being drawn to the subtask and losing sight of the overall goal, forgetting to check whether current actions are on track toward goal attainment, or having difficulty sequencing and switching attention between different subtasks relevant to goal attainment. It would be valuable to be able to distinguish among all of these component processes underlying goal-directed behavior and other possible impediments to goal attainment and to determine which cognitive processes contribute most to functional difficulties. In particular, given the body of research suggesting that selective attention and working memory play a central role in goal-directed functioning,[29] we wanted to determine to what extent working memory ability (as assessed by neuropsychological measures) predicts performance on a complex unstructured functional task.

We set out to address challenges in parsing the contributions of various component processes in reliable observational ratings in ecologically valid settings. We developed an assessment tool that involves both a challenging task that engages executive functions and a rating system to quantify observations in specific cognitive domains. The GPS evaluates a participant's ability to plan and execute a complex task commonly encountered in everyday life, requiring researching options and making a decision based on information gathered in a limited time while following specified rules.[30] The participant's functional ability to guide attention and related aspects of behavior based on goals is evaluated using a rating scale for overall performance and 8 subdomains of executive functioning: Planning, Initiation, Maintenance of Attention, Attentional Sequencing and Switching of between task subcomponents, Self-monitoring, Flexible Problem Solving, Task Execution, and Learning and Memory. To assist in rating of performance and to ensure rating consistency, the domains evaluated were operationally defined in rating forms and detailed calibration instructions were provided in an instruction manual. We optimized the calibration and operationalization of the rating system and then tested the interrater reliability.

We report the results of a study testing the GPS with 19 participants with chronic (6+ months) acquired brain injury. We evaluated the extent to which the GPS overall score reflects functional performance that is captured by an existing validated but more general functional assessment, the MET. We further assessed the extent to which participants' performance on neuropsychological measures assessing attention/executive function and memory abilities explains variance in their performance on a complex unstructured functional task such as the GPS.

METHODS

Study participants

Nineteen participants (mean age = 41.4 years, SD = 12.9; mean education = 14.7 years, SD = 1.9) with a history of chronic acquired brain injury (range = 6 months to 39 years postinjury) diagnosed by a neurologist and or physiatrist specializing in brain injury were included in the study. Eighteen participants had a history of traumatic brain injury with loss of consciousness (ranging from mild to severe), and 1 participant had a history of stroke. All were independent in performing basic ADL but reported mild to moderate difficulties on daily tasks involving organization, problem solving, multitasking, and sustained attention. At the time of assessment, 10 participants were not working, 5 were going to school, and 4 were gainfully employed but indicated difficulties with performing work and school tasks. All participants were on a stable medication regimen and had no active illicit drug use, aphasia, or other conditions that would impede participation in the assessments. Before the injury, all were gainfully employed. Additional details describing each participant's brain injury severity, work/school status, primary functional complaint at the time of assessment, and neuroimaging findings are included in the Supplemental Digital Content Table 1 (available at http://links.lww.com/JHTR/A65). Participants were referred to the investigators from their physicians and treatment providers at several San Francisco Bay Area hospitals. This study was approved by institutional review boards at participating institutions including the Veteran's Administration Medical Centers in San Francisco and Martinez, University of California, San Francisco and Berkeley, and California Pacific Medical Center. All participants provided informed consent prior to any study procedures taking place.

Measures

Participants were evaluated with a battery of tests consisting of functional and neuropsychological measures.

The Goal Processing Scale

The GPS assessment involves 2 components: a challenging task that engages executive control, and a rating system to quantify observations. Participants are instructed to plan and execute a task requiring them to gather and compare information about 3 different activities (or products/services–as designated on alternate forms) of their choice, using the available means while following specified rules in a limited time (30 minutes). Participants work in an office equipped with a computer with Internet access, a telephone, yellow pages telephone book, blank paper, pen, calculator, and a clock. They are given the task instructions page (see Table 1), which is also read aloud by the evaluator who is present in the room during the evaluation. To ensure participants' understanding of the task, the evaluator asks them to reiterate the instructions and to request clarifications as needed. After ensuring that participants understand the instructions, during the planning phase they are asked to decide on the actual goals and parameters of the task and to identify the steps needed to complete the task. During the 30-minute task execution stage, participants

TABLE 1 *Goal Processing Scale task instruction*

We would like to see how you work on a task

First, you will be asked to pick a general activity area from the choices below:
- Your ultimate vacation
- Weekend getaway
- New hobby or leisure activity
- Sporting or entertainment event
- Fitness program

Then, you will compare 3 different options within the activity area you selected, based on the following criteria: cost, convenience, customer review, and a criterion of your choice.

For example, if you select a *weekend getaway* as a general activity area, you may compare 3 possible destinations (such as Tahoe, Napa, and Monterey) using the following criteria: cost, convenience of location, customer review, and the criteria of your choice—family friendly activities.

In the planning phase, please tell the evaluator

Your choice of an activity area, and 3 possible options within that category you will compare

The 4 comparison criteria:
- A criteria of your choice
- The following 3 criteria: (1) cost; (2) convenience; and (3) customer review

List steps involved in completing this task

During the task

At 8 and 20 min into the task, you must stop what you are doing, check, and review your work. Tell the evaluator:
- a. At what stage you are at in completing the task relative to outlined goals
- b. If you are satisfied with your work
- c. If you have made any errors: possible errors include breaking task rules, or going off task (eg, spending time on activities unrelated to task, or changing task goals once you start)
- d. You will have 1 min to ask questions. Questions will only be answered during the breaks.

At the end of the task

Give 3 reasons why you would choose one of the activities over the others.

Tell the evaluator how you feel about your work, and if you would change anything if you were to do this again.

Task rules

Please do not select the exact same choices as in the example provided.

You have 30 min to complete the task (after completing the planning phase).

You cannot use the same method (ie, the Internet or phone) to find information about more than 2 of the selected activities being researched.

If you are using a phone, it needs to be on speaker phone

Once you start the task, you will not be able to ask any questions until the break.

There may be unexpected interruptions or distractions during the task, if this happens, please continue the task using the available means.

You are responsible for tracking time during the task, self-initiating the breaks at 8 and 20 min into the task, and stopping at 30 min for a final review.

are evaluated on their ability to effectively execute the task on the basis of their identified plan and adherence to task rules.

The areas evaluated during the *planning stage* include the ability to understand task instructions and ask for clarifications when needed; to decide on and identify realistic goal(s); and to organize and prioritize steps involved in actual task execution. After the task goals and plan are decided upon, the participants are told to execute the task on the basis of their identified plan and task rules. The domains assessed during the *task execution stage* include the ability to *initiate* task-directed activities; *maintain attention* on a task both in a nondistracting environment and during the built-in task distractions; *self-monitor* performance (including inhibiting task activities to stop at specified times, review performance, notice, and correct errors); *sequence and switch attention* between and among the identified task subcomponents; demonstrate *flexibility in approaching alternate solutions* when the situation changes (eg, the ability to continue with specified task goals when the preferred means of obtaining information such as the Internet or phone become unavailable); *memory*–including both the ability to recall strategies when needed and the ability to correct previously noted errors; and actual *execution*, reflecting the accuracy of completion of identified task goals and effectiveness of time management while executing steps relevant to the identified plan and goals. Awareness of one's ability, reflected in the degree of mismatch between the performance that a participant predicts and actual performance, is also evaluated.

Functional performance in these domains is rated on a scale: 0 (*not able*) to 10 (*absolutely not a problem*). The GPS overall performance score is the average of the 8 subdomain scores. To assist in rating of performance on the subdomains and to ensure rating consistency, the domains were operationally defined (see Table 2). Furthermore, a GPS Rating Instruction Manual operationally defines and calibrates the following: (1) the cognitive domains evaluated; (2) the task-based context; (3) the rating scale; and (4) the objective criterion-based scoring used for evaluation. Specifically, in developing operational definitions for cognitive domain components, we created systematic and objective end points that can be evaluated consistently by independent raters, thus increasing interrater reliability. First, in the Rating Instruction Manual, each subdomain is operationally defined to clarify the concept that is to be evaluated. Second, a further qualitative description provides a contextual understanding of how the qualities of the subdomain are likely to be observed during an actual task performance; this creates preparatory set and framework and cues the evaluator to attend to specific elements of the task. Third, a task-specific rating scale is defined with qualitative descriptions of each numeric end point. Finally, a scoring grid composed of objective task-based criteria is provided for rating by the evaluator.

TABLE 2 *Goal Processing Scale domains rated*

1. Planning
   a. Understanding task instructions and requesting clarifications when needed
   b. Identifying the areas of interest
   c. Deciding on a goal (task)
   d. Breaking down the task into steps involved
2. Initiation
   a. Initiation of activity on a task
3. Self-monitoring
   a. Periodically stopping (inhibiting) ongoing activity
   b. Reassessing task performance
   c. Noticing errors
   d. Correcting errors
4. Maintenance of attention
   a. Sustaining attention on a task in a nondistracting environment
   b. Sustaining attention on a task in a distracting environment
5. Sequencing and switching of attention
   a. Switching attention between task subcomponents
6. Flexible problem solving
   a. Flexibility in approaching alternate solutions
7. Task execution
   a. Effectiveness in executing steps relevant to the identified plan and goals
   b. Effective time management
   c. Accurate task completion
8. Learning and memory
   a. Ability to recall strategies (or steps involved) when needed
   b. Ability to learn from mistakes

Overall score (average of domains) = 1-8

Evaluators, professionals experienced in assessing individuals with brain injury (an occupational therapist, speech therapist, and neuropsychology fellow), were trained in administering the GPS and in rating the performance using the manual. On average, during the training, each evaluator completed 4 to 5 GPS evaluations before becoming proficient in administration and scoring. To evaluate interrater reliability of the GPS, 2 experienced raters directly observed and independently rated each participant's task performance.

The Multiple Errands Test

The MET is an unstructured functional task that allows clinicians to directly assess a patient's ability to follow outlined rules and complete multiple "real-world" tasks in a limited time period. Keeping the main subcomponents of the task outlined in Alderman et al[23] and Knight et al,[24] we adapted the MET for use in local hospital settings and we developed 3 alternate forms to allow for longitudinal assessments. Both tests were administered to participants on the same day, with MET testing preceding testing with the GPS.

Participants were given written instructions and a hospital map and asked to complete 12 subtasks in 40 minutes while following 11 specified task rules. Evaluators followed participants while they were performing the tasks, directly observing their actions. The 12 subtasks involved performing different activities (eg, buying specified items, collecting an envelope with instructions from a designated location, using a hospital phone system to reach a particular person), obtaining information (eg, determining the opening time of the hospital shop on a Saturday or the US city predicted to reach the highest temperature tomorrow), and stopping ongoing activity to meet the evaluator at a specified place and time. Task performances were rated as 0 (*successful completion*), 1 (*partial completion*), and 2 (*failure*) so that a maximum number of task failure points are 24. The task rules included: not leaving the hospital grounds; spending a maximum of $5.00 (they are given a $10.00 bill); not entering a building or room without completing part of the task inside; not going back to the same area more than once; buying no more than 2 items from 1 location; not talking about things unrelated to the task; completing a task in the allotted time; not using personal money or cell phone; and giving the evaluator test materials at the end of the task.

Neuropsychological assessment

The neuropsychological battery used in this study was developed to assess performance in cognitive domains of complex attention/executive function and memory that are commonly affected by traumatic brain injury. To prevent participants' fatigue, the neuropsychological battery was administered on a day different from that of the GPS and the MET, in a majority of cases, within the same week. All neuropsychological test data were scored on the basis of standardized age and (when available) educational norms and transformed into *z* scores for consistency. Individual neuropsychological test *z* scores were averaged into the cognitive domain scores as follows:

Working Memory was assessed with (1) Auditory Consonant Trigrams,[31] requiring recall of 3 consonants after counting backwards by 3 from a specified number for 9, 18, and 36 seconds, and (2) Letter Number Sequencing, Wechsler Adult Intelligence Scale–III,[32] requiring mental reordering of scrambled letter-number series of increasing lengths.

Sustained Attention was assessed with the time and error score on the Digit Vigilance Test,[33] requiring detection and crossing out of a specified number intermixed with other numbers on 2 pages.

Inhibition of automatic responding was assessed with the time and error score on the Stroop Inhibition Delis Kaplan Executive Function System (DKEFS),[34] in which words are printed in contrasting ink colors, and participants are instructed to name the color of the ink instead of the more automatic response of reading the word.

Mental Flexibility was assessed with the following: (1) Trails B,[33] requiring rapid tracing between letters and numbers to connect them in alternating order; (2) Design Fluency-Switching DKEFS,[34] requiring alternating between empty and filled dots while generating different designs with 4 lines; (3) Verbal Fluency-Switching DKEFS,[34] requiring generating words that belong to 2 specified categories and alternating between them; and (4) Stroop Inhibition-Switching DKEFS,[34] during which the participant is presented with words printed in contrasting ink colors, some of which are contained in boxes; they are instructed to name the color of the ink, unless the word is inside the box, in which case they are to read the word.

Generative Ability was assessed with DKEFS Verbal Fluency Total Correct (letter and category fluency) and Design Fluency Total Correct (empty and filled dots).

Learning/Memory was assessed with the Hopkins Verbal Learning Test–Revised,[35] requiring participants to learn 12 words after 3 learning trials and to recall them after 25 minutes; and with Brief Visual Memory Test–Revised,[36] requiring participants to learn and reproduce 6 abstract designs over 3 learning trials and to reproduce them after 25 minutes.

Statistical analysis

Interrater reliability for the GPS

For an assessment tool such as the GPS that assesses complex "real-life" performance, achieving interrater consistency is both crucial and a challenge. This was a major consideration in operationalizing the GPS subdomain ratings. To determine whether the GPS reliably operationalizes the subdomains of executive functioning it is designed to measure, each participant's performance was independently rated by 2 trained raters using the GPS Rating Instruction Manual. Intraclass correlation coefficients (ICCs) were then calculated for each subdomain and for the overall GPS score, which was derived by averaging across all subdomain scores. The ICC measures the similarity between raters' scores for the individual subjects. For the purposes of these analyses, we assumed that rater was a fixed effect, ICC (3,1) as described by Shrout and Fleiss, and calculated ICC measurements accordingly. That is, our raters were assumed to be the entire population of raters. However, results were virtually identical to those assuming rater to be a random effect, that is, raters assumed to be a random subset of all possible raters, ICC (2,1).[37] All ICCs were calculated using a SAS macro, intracc.sas.[38]

Correlation with an established functional assessment test

To determine whether the participants' performance on the GPS correlated with MET performance, we first used Spearman correlation coefficients to compare MET summary scores to the overall GPS score (the average of the subdomain scores). Nonparametric tests were used to account for the nonnormality of the MET data, which consist of counts of the number of task rule breaks and task failures. To examine the extent to which GPS subdomains might indicate the sources of reduced functional performance measured by the more general MET, we then performed correlation analysis on the MET summary scores with each of the individual GPS subdomains.

Although we have examined a number of subdomains, we report nominal *P* values without adjustment for multiple testing. Such adjustment would be focused on avoidance of 1 or more results, with $P < .05$ in the case where all differences are truly zero,[39–41] which is an unrealistic hypothesis about the state of nature in this situation. In addition, adjustment would require that each result detract from the others, but there are clear relations between the domains that we are examining, and these permit coherent sets of findings to reinforce each other rather than detract from one another.

Correlation between GPS subdomains and neuropsychological measures

To determine whether participants' performance on neuropsychological tests assessing attention, executive, and memory function relates to and predicts their performance on a complex functional task (as assessed by GPS overall and subdomain performance), we examined correlations between the GPS and the neuropsychological test domain scores. We further examined which, if any, of these cognitive abilities predicts a participant's performance on a complex unstructured functional assessment such as the GPS.

Pearson correlation coefficients were used to determine the degree of association between attention/executive functioning and memory abilities (as assessed by neuropsychological test performance domain scores) and functional performance in low-structure setting (as assessed by GPS overall and subdomain scores). For those relations where we observed significant correlation, we used linear regression models to further examine and quantify these associations. All statistical analyses were conducted using SAS v9.2.

RESULTS

Interrater reliability

Intraclass correlation coefficients assessing interrater agreement demonstrated a very high correlation of 0.97 for GPS overall performance score. Seven of the 8 GPS subdomain scores demonstrated a correlation of 0.80 or higher (see Figure 1), suggesting good interrater reliability.

GPS and functional measures

MET and overall GPS functional performance

To determine whether the GPS successfully captures functional performance on an independent complex functional task, we compared it with the MET using Spearman correlation coefficients. We observed a correlation of −0.59 between the GPS and the MET task failures ($P = .012$) and a correlation of −0.60 between the GPS and MET rule breaks ($P = .01$).

Relations between GPS subdomains and MET performance

To determine which GPS subdomains most strongly correlated with MET scores, Spearman correlation coefficients were calculated. The performance on the GPS Task Execution and Learning and Memory subdomains significantly correlated with lower number of both rule breaks and task failures on the MET. In addition, better performance on the GPS Attentional Sequencing and Switching significantly correlated with a lower number of rule breaks on the MET and better performance on the GPS Self -monitoring domain significantly correlated with a lower number of task failures on the MET. The remaining subdomains were not significantly correlated with either the rule breaks or the task failures on the MET. The fact that some subdomains correlate with the MET more strongly than others suggests that GPS may be valuable for determining the *sources* of reduced functional performance that are measured more generally with a test such as the MET (see Table 3).

GPS and neuropsychological measures

To determine whether participants' performance on neuropsychological tests assessing attention, executive, and memory abilities relates to and predicts their performance on a complex functional task (as assessed by GPS overall and subdomain performance), we examined correlations between (1) overall and the 8 domain scores of the GPS and (2) the 6 neuropsychological test domain scores. Therefore, 54 pairwise correlations were generated (see Supplemental Digital Content Table 2, available at http://links.lww.com/JHTR/A66).

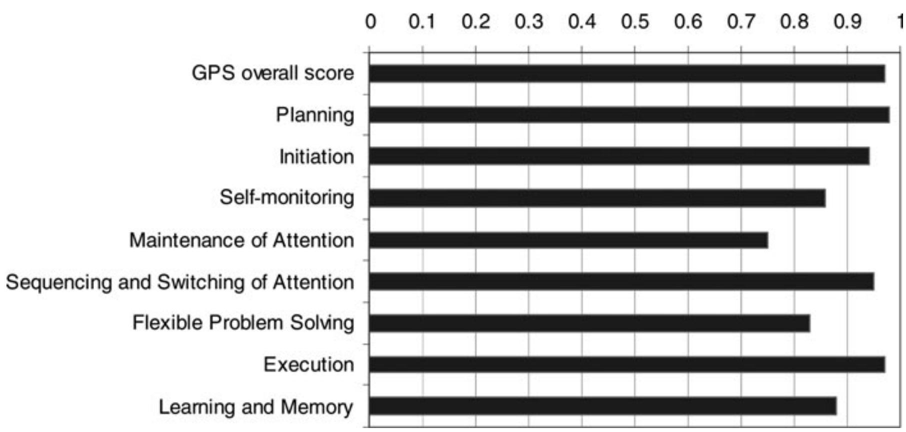

Figure 1. Interrater reliability (intraclass correlation coefficient) for Goal Processing Scale (GPS) overall and subdomain scores.

TABLE 3 *Relations between performance on the GPS and the MET*

	MET Measures			
	Number of rule breaks		Number of task failures	
GPS Domain	Spearman *r*	*P*	Spearman *r*	*P*
Attentional Sequencing and Switching	−0.70	**.002****	−0.35	.16
Task Execution	−0.66	**.004****	−0.64	**.005****
Self-Monitoring	−0.45	.07	−0.47	**.05***
Learning and Memory	−0.57	**.02***	−0.60	**.01****
Flexible Problem Solving	−0.21	.44	−0.38	.15
Planning	−0.41	.09	−0.46	.06
Initiation	−0.27	.28	−0.13	.62
Maintenance of Attention	−0.23	.36	−0.24	.35

Abbreviations: GPS, Goal Processing Scale; MET, Multiple Errands Test.
*$P < .05$.
** $P < .01$.

Performance on the neuropsychological measures assessing working memory significantly correlated with performance on the GPS overall ($r = 0.66$, $P = .005$) as well as with the GPS subdomains assessing Self-monitoring ($r = 0.55$, $P = .02$), Attentional Sequencing and Switching ($r = 0.70$, $P = .002$), Task Execution ($r = 0.66$, $P = .005$), and Learning and Memory ($r = 0.61$, $P = .02$).

Performance on the neuropsychological measures assessing learning and memory abilities significantly correlated with performance on the GPS overall ($r = 0.61$, $P = .01$) as well as with the GPS subdomains assessing Self-monitoring ($r = 0.59$, $P = .02$), Task Execution ($r = 0.56$, $P = .023$), Learning and Memory ($r = 0.46$, $P = .02$), and Flexible Problem Solving ($r = 0.55$, $P = .04$).

The remaining neuropsychological measures assessing sustained attention, inhibition, mental flexibility, and generative ability did not significantly correlate with any of the GPS subdomain scores.

Because correlations are bidirectional, we used univariate regression models to further examine the significant associations between performance on the GPS and the neuropsychological measures of working memory and learning and memory. This was done both to quantify the effects of working memory and learning/memory abilities on GPS complex functional task performance and to estimate confidence intervals for these effects.

Results of regression models examining the potential association between working memory ability and GPS are presented in Table 4. For each increase of 1 SD in working memory (raw scores having been converted to *z* scores), overall GPS performance increased an average of 1.16 points (on a scale ranging from 0 to 10). Working memory had the largest effect on the GPS Attentional Sequencing and Switching, Task Execution, and Learning and Memory subdomains; for each increase of 1 SD in the neuropsychological working memory score,

TABLE 4 *Relations between working memory domain and GPS performance*

GPS Domain	Coefficient	95% CI	*P*
Overall performance	1.16	0.40-1.91	**.006****
Attentional Sequencing and Switching	2.09	0.87-3.30	**.003****
Task Execution	2.04	0.72-3.36	**.005****
Self-Monitoring	1.38	0.19-2.57	**.03***
Learning and Memory	2.30	0.61-4.01	**.01****
Flexible Problem Solving	1.14	−0.47 to 2.75	.15
Planning	0.69	−0.76 to 2.13	.33
Initiation	0.17	−0.10 to 0.45	.19
Maintenance of Attention	0.26	−0.53 to 1.05	.49

Abbreviation: GPS, Goal Processing Scale.
*$P < .05$.
** $P < .01$.

the GPS functional performance in these subdomains increased by an average of more than 2 points.

The overall correlation between the GPS and learning/memory was not as strong as that with working memory; each increase of 1 SD unit in learning/memory resulted in an average increase in the overall GPS score of 0.63 points (see Table 5). This same increase resulted in an approximately 1-point increase in the GPS Learning and Memory, Self-monitoring, Task Execution, and Attentional Sequencing and Switching subdomains.

DISCUSSION

Results of the this study suggest that the GPS may be a useful tool for assessing the contribution of different subdomains of cognitive processing to functional performance on complex tasks in low-structure, ecologically valid settings. Several challenges were addressed in its development and validation. In any assessment of complex "real-life" performance, achieving interrater consistency is both crucial and a challenge. This was a major consideration in designing 2 key components of the GPS tool: the complex functional task and the system for rating cognitive processing during the complexities of individual performance. Specifically, the GPS was designed to evaluate participants' ability to plan and execute an activity commonly encountered in everyday life, requiring researching options and making a decision on the basis of information gathered in a limited time, while following specified rules. To assist in rating of performance and to ensure rating consistency, the executive function subdomains evaluated were operationally defined and detailed calibration instructions were outlined in a GPS Rating Instruction Manual.

In this study, the interrater agreement for 2 trained independent raters was high both for the GPS overall performance and for the 8 executive function performance subdomain scores. This suggests that the systematized and calibrated rating system outlined in the GPS Rating Manual was sufficient to achieve high interrater reliability. It is important to note that GPS evaluators were clinicians (occupational, speech therapists, and neuropsychology fellows) experienced in working with individuals with brain injury.

As a gross validation of the GPS in quantifying functional performance in a complex, low-structure setting, we evaluated the extent to which the GPS overall score reflects performance that is captured by an existing validated but more general functional assessment, the MET. The higher GPS overall performance score correlated well with fewer task failures and rule breaks on the MET, suggesting good convergent validity between the 2 functional assessment tools. Furthermore, among GPS subdomains, better overall Task Execution and Learning and Memory correlated with better performance on the MET (lower number of both rule breaks and task failures). Better performance on the GPS subdomain Attentional Sequencing and Switching between different task components was associated with fewer rule breaks on the MET, possibly reflecting one's ability to switch between both task rules and task demands while performing the designated activity.

Having established basic aspects of interrater reliability and validity, we then turned to evaluating which cognitive domains may contribute the most to goal-based performance on complex functional tasks. As outlined by Duncan[28] and others,[26,27] individuals with executive dysfunction can frequently plan and verbalize their goals and subtasks but often fail to maintain goal-directed behavior during execution. Goal neglect, or a failure in maintaining the goal, can occur as a consequence of being distracted from a goal including being drawn to the subtask and losing sight of the overall goal, forgetting to check whether current actions are on track toward goal attainment or having difficulty sequencing and switching attention between different subtasks relevant to goal attainment. It would be valuable to be able to distinguish among all of these component processes underlying

TABLE 5 *Relations between learning/memory domain and GPS performance*

GPS Domain	Coefficient	95% CI	*P*
Overall performance	0.63	0.03-1.23	**.04***
Attentional Sequencing and Switching	1.10	0.09-2.11	**.03***
Task Execution	1.16	0.13-2.20	**.03***
Self-Monitoring	0.86	−0.01 to 1.72	**.05***
Learning and Memory	1.24	−0.11 to 2.55	.07
Flexible Problem Solving	0.95	−0.26 to 2.17	.11
Planning	0.57	−0.42 to 1.56	.24
Initiation	0.08	−0.12 to 0.28	.39
Maintenance of Attention	0.16	−0.39 to 0.71	.54

Abbreviation: GPS, Goal Processing Scale.
*$P < .05$.

goal-directed behavior and other possible sources of poor goal attainment and to determine which cognitive processes contribute most to functional difficulties.

There is a significant gap in knowledge of the cognitive subprocesses that influence and/or determine real-life functional performance in patients with brain injuries. One line of thought is that the integrity of the information maintained in working memory depends on the effectiveness of selection attention and can be disrupted by the entry of nonrelevant information. Thus, the pathway between goals and actions may depend on the effectiveness of selective attention and working memory and may be vulnerable to distracting processes. Selective attention and working memory have been linked to prefrontal cortex and its interactions with different brain regions,[42–45] and are thought to play a central role in goal-directed functioning.[29]

We performed a preliminary test of the proposition that working memory plays a central role in effective goal management in complex, ecologically valid situations for patients with acquired brain injury. The results of this study showed that working memory ability in the context of distractions (as assessed by independent neuropsychological measures) predicts overall performance on a complex unstructured functional task. Furthermore, an individual's working memory performance had the greatest predictive value for GPS functional performance in the Attentional Sequencing and Switching, Task Execution, and Learning and Memory subdomains, all domains that functionally should depend on working memory.

Participants' performances on neuropsychological domains assessing sustained attention, mental flexibility, inhibition, and generative ability did not correlate with their functional performance on the GPS. One explanation may be that the neuropsychological measures and the GPS assess somewhat different abilities on the basis of performance in somewhat different contexts. For example, there is no measure of sustained attention during the entire neuropsychological evaluation. The Digit Vigilance Test used to assess sustained attention is a highly structured task administered in a quiet structured environment. The score is based on completion time and number of errors. In contrast, the GPS Maintenance of Attention subdomain assesses participant's performance for the entire 30-minute long GPS assessment in both distracting and nondistracting environments. Similarly, mental flexibility in the setting of unexpected obstacles toward accomplishing a functional goal was not assessed by the neuropsychological test we used. Overall, these differences highlight the potential value and importance of observing functional performance in ecologically valid settings.

Although the approach embodied by the GPS may fill an important gap in the field in assessing and quantifying subcomponents of functional task performance, there are a number of limitations. One limitation of this study is the small number of patients; our investigation needs to be replicated on a larger group. This scale primarily addresses aspects of executive function associated with dorsolateral prefrontal cortex. Development of additional tools assessing behavioral dyscontrol and disinhibition would be worthwhile. The test-retest reliability of the GPS was not addressed in this study. We have developed and are currently evaluating alternate forms of the GPS to allow for the assessment of the effects of therapeutic interventions on different subdomains of executive functioning.

The GPS may also be a useful tool for investigations of the neural underpinnings of functional performance in real-world settings. Moreover, some approaches used in the GPS, such as the process for observed ratings, may be more generally applicable for assessing contributors to functional performance in other settings that may be relevant to an individual's personal occupation or other goals. Finally, improved ability to define each individual's strengths and weaknesses on subcomponents of executive function in complex real-life environments may also help in the development and evaluation of the efficacy of therapeutic interventions.

REFERENCES

1. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4th ed. New York, NY: Oxford University Press; 2004.
2. Hecaen H, MLA . *Human Neuropsychology*. New York, NY: John Wiley & Sons; 1978.
3. Vanderploeg RD, Curtiss G, Belanger HG. Long-term neuropsychological outcomes following mild traumatic brain injury. *J Int Neuropsychol Soc*. 2005;11(3):228–236.
4. Belanger HG, Curtiss G, Demery JA, et al. Factors moderating neuropsychological outcomes following mild traumatic brain injury: a meta-analysis. *J Int Neuropsychol Soc*. 2005;11(3):215–227.
5. Binder LM, Rohling ML, Larrabee GJ. A review of mild head trauma, part I: meta-analytic review of neuropsychological studies. *J Clin Exp Neuropsychol*. 1997;19(3):421–431.
6. Frencham KA, Fox AM, Maybery MT. Neuropsychological studies of mild traumatic brain injury: a meta-analytic review of research since 1995. *J Clin Exp Neuropsychol*. 2005;27(3):334–351.
7. Cicerone KD, Azulay J. Diagnostic utility of attention measures in postconcussion syndrome. *Clin Neuropsychol*. 2002;16(3):280–289.
8. Mathias JL, Beall JA, Bigler ED. Neuropsychological and information processing deficits following mild traumatic brain injury. *J Int Neuropsychol Soc*. 2004;10(2):286–297.
9. Mathias JL, Coats JL. Emotional and cognitive sequelae to mild traumatic brain injury. *J Clin Exp Neuropsychol*. 1999;21(2):200–215.
10. Ownsworth T, McKenna K. Investigation of factors related to employment outcome following traumatic brain injury: a critical

review and conceptual model. *Disabil Rehabil*. 2004;26(13):765–783.

11. Drake AI, Gray N, Yoder S, et al. Factors predicting return to work following mild traumatic brain injury: a discriminant analysis. *J Head Trauma Rehabil*. 2000;15(5):1103–1112.
12. Machamer J, Temkin N, Fraser R, et al. Stability of employment after traumatic brain injury. *J Int Neuropsychol Soc*. 2005;11(7):807–816.
13. Doctor JN, Castro J, Temkin NR, et al. Workers' risk of unemployment after traumatic brain injury: a normed comparison. *J Int Neuropsychol Soc*. 2005;11(6):747–752.
14. Alexander M, Stuss DT. Frontal injury: impairments of fundamental processes lead to functional consequences. *J Int Neuropsychol Soc*. 2006;12(2):192–193.
15. Burgess PW, Alderman N, Forbes C, et al. The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *J Int Neuropsychol Soc*. 2006;12(2):194–209.
16. Evans J. Can executive impairments be effectively treated? In: *Effectiveness of Rehabilitation for Cognitive Deficits*. New York, NY: Oxford University Press; 2005:247–257.
17. Chan RCK, Shum D, Toulopoulou T, Chen EYH. Assessment of executive functions: review of instruments and identification of critical issues. *Arch Clin Neuropsychol*. 2008;23:201–216.
18. Marcotte TD, Scott JC, Kamat R., Heaton RK Neuropsychology and the prediction of everyday functioning. In:Marcotte TD, Grant I, eds. *Neuropsychology of Everyday Functioning*. New York, NY: The Guilford Press; 2010: 5–38.
19. Chelune GJ, Heaton RK, Lehman RAW. Neuropsychological and personality correlates of patients complaints of disability. In: Goldstein G, Tarter RE, eds. *Advances in Clinical Neuropsychology*. New York, NY: Plenum Press; 1986:95–126.
20. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: review of the literature on everyday cognitive skills. *Neuropsychol Rev*. 2003;13(4):181–197.
21. Anderten Loeb P. *Independent Living Scales*. San Antonio, TX: The Psychological Corporation; 1996.
22. Schwartz MF, Segal M, Veramonti T, Ferraro M, Buxbaum LJ. The Naturalistic Action Test: a standardized assessment for everyday action impairment. *Neuropsychol Rehabil*. 2002;12(4):311–339.
23. Alderman N, Burgess PW, Knight C, et al. Ecological validity of a simplified version of the Multiple Errands Shopping Test. *J Int Neuropsychol Soc*. 2003;9(1):31–44.
24. Knight C, Alderman N, Burgess P. Development of a simplified version of the Multiple Errands Test for use in hospital settings. *Neuropsychol Rehabil*. 2002;12:231–255.
25. Dawson DR, Anderson ND, Burgess P, Cooper E, Krpan KM, Stuss DT. Further development of the Multiple Errands Test: standardized scoring, reliability and ecological validity for the Baycrest version. *Arch Physical Med Rehabil*. 2009;90(11):S41–S51.
26. Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. *Brain*. 1991;114:727–741.
27. Wilson B. Neuropsychological rehabilitation. *Annu Rev Clin Psychol*. 2008;4:141–162.
28. Duncan J. Disorganization of behavior after frontal lobe damage. *Cognitive Neuropsychol*. 1986;3:271–290.
29. Fuster JM. The prefrontal cortex–an update: time is of the essence. *Neuron*. 2001;30(2):319–333.
30. Novakovic-Agopian T, Chen A, Rome S, et al. Assessment of executive functioning with goal processing scale–a complex "real-life" functional performance test: pilot results. *J Int Neuropsychol Soc*. 2010;16(suppl 1):126.
31. Stuss D, Stethem LL, Pelchat G. Three tests of attention and rapid information processing: an extension. *Clin Neuropsychol*. 1988;1:139–152.
32. Wechsler D. *Wechsler Adult Intelligence Scale–III Manual*. San Antonio, TX: The Psychological Corporation; 1997.
33. Heaton R, Miller W, Taylor MJ, Grant I. *Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographically Adjusted Neuropsychological Norms for African American and Caucasian Adults*. Lutz, FL: Psychological Assessment Resources; 2004.
34. Delis D, Kaplan E, Kramer J. *Delis-Kaplan Executive Function System*. San Antonio, TX: Pearson Assessment and Information; 2001.
35. Brandt J, Benedict R. *Hopkins Verbal Learning Test-Revised. Professional Manual*. Odessa, FL: Psychological Assessment Resources; 2001.
36. Benedict R. *Brief Visuospatial Memory Test-Revised*. Odessa, FL: Psychological Assessment Resources; 1997.
37. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
38. Hamer RM. *SAS Macro: intracc.sas*. Cary, NC: SAS Institute Inc; 1991.
39. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1(1):43–46.
40. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *Am J Epidemiol*. 1995;142(9):904–908.
41. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316(7139):1236–1238.
42. D'Esposito M, Detre JA, Alsop DC, et al. The neural basis of the central executive system of working memory. *Nature*. 1995;378(6554):279–281.
43. Kane MJ, Engle RW. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon Bull Rev*. 2002;9(4):637–71.
44. Chen AJW, Novakovic-Agopian T, Nycum TJ, et al. Training of goal-directed attention regulation enhances control over neural processing for individuals with brain injury. *Brain*. 2011;134:1541–1554.
45. Novakovic-Agopian T, Chen AJW, Rome S, et al. Rehabilitation of executive functioning with training in attention regulation applied to individually defined goals: a pilot study bridging theory, assessment, and treatment. *J Head Trauma Rehabil*. 2011;26(5):325–338.