Kettle Test—A Brief Measure of Cognitive Functional Performance: Reliability and Validity in Stroke Rehabilitation

Adina Hartman-Maeir, Hagit Harel, Noomi Katz

OBJECTIVES. We examined the reliability and validity of the Kettle Test, a brief performance measure based on a complex everyday task designed to tap into basic and higher level cognitive processes.

METHOD. Participants included 21 people attending stroke rehabilitation and 4 occupational therapists for the reliability analysis, 36 people at discharge from stroke rehabilitation, and 36 age-matched healthy control participants for the validity analyses. Instruments included a battery of conventional cognitive measures and functional outcomes.

RESULTS. Interrater reliability was found to be high. Stroke survivors at discharge from rehabilitation were found to require significantly more assistance on the Kettle Test than control participants ($p < .000$); their scores on the Kettle Test were significantly and moderately correlated with the conventional cognitive and functional outcome measures.

CONCLUSIONS. The results support the reliability and validity of the Kettle Test as a top-down measure of cognition-in-function in people at discharge from stroke rehabilitation.


People with acquired brain injury are at high risk for cognitive impairments that have a detrimental impact on occupational performance (Cicerone et al., 2005; Donnovan et al., 2008; Hershkovitz & Brill, 2007; Zinn et al., 2004). Occupational therapy has developed a unique body of knowledge regarding the relationship between cognition and occupation involving occupation-based, client-centered intervention (Gillen, 2009; Katz, 2005). Conventional “tabletop” measures of cognition are valuable diagnostically but have limited ecological validity and do not fully address the functional implications of cognitive deficits. Moreover, the interdisciplinary rehabilitation community has emphasized the need for ecologically valid measures of cognition in function (Burgess et al., 2006). This current emphasis is an important milestone in the conceptualization of measurement in health-related professions. Instruments that follow these directives are essential and are integral to the occupational therapy philosophy and practice (American Occupational Therapy Association, 2008). Occupational therapy is in a unique position to fulfill a central role in this development and is establishing its expertise in the development of top-down assessments (Hartman-Maeir, Katz, & Baum, 2009; Law, Baum, & Dunn, 2005).

Evaluation of outcomes in stroke rehabilitation, beyond basic activities of daily living (BADLs), has become extremely important as more people survive stroke and return to their homes in the community. One primary concern at discharge from rehabilitation is predicting the ability to function independently in the community or the degree of assistance required to do so. Instrumental activities of daily living...
Development of the Kettle Test

The Kettle Test (Hartman-Maeir et al., 2005) was developed to provide a brief performance-based assessment of an IADL task that can be conducted in a clinical setting or at home. The test was designed to tap into a broad range of cognitive skills within a functional context to assist in the clinical decision-making process regarding the need for assistance in daily living skills of adults with suspected cognitive disabilities. The basic task of preparing a hot beverage was chosen because of its functional significance, broad cultural relevance, and feasibility (in terms of tools and time requirements) in multiple settings (e.g., clinic, home). The task complexity was elevated to enable the assessment of basic and higher-level cognitive–functional skills that are required for IADLs, with the following elaborations:

1. The essential task of preparing oneself a hot beverage was expanded to include preparing an additional cup of hot beverage for the therapist; the additional cup differs in two ingredients from that requested by the client, providing additional load on working memory.
2. The electric kettle is emptied and disassembled (lid and electric cable are disconnected from the body of the kettle) to challenge problem-solving skills and safety judgment regarding the use of electricity.
3. Additional kitchen utensils and ingredients are placed in the immediate task environment (on a tray) as distracters to increase attention demands.

This design creates a brief task using familiar objects and settings that are presented in a manner that targets cognitive skills underlying complex ADLs. The observation is structured so that the rater is required to score the performance on 13 discrete steps of the task (e.g., turning on the faucet, filling the kettle with 2 cups of water). Clear guidelines for cueing are provided, and the rater scores each step according to the degree of cueing that was necessary to complete the step (0 through 4). Total scores range from 0 to 52 (higher scores indicate more assistance) and can be transformed to clinically meaningful categories of independence on the test (independent, mild assistance, considerable assistance).

Initial research to examine convergent and ecological validity was conducted on the Kettle Test in a sample of 41 elderly clients (mean age = 75.2) referred to a geriatric assessment clinic because of suspected cognitive disabilities (Hartman-Maeir, Katz, & Armon, 2004). Small to moderate significant correlations were found between Kettle Test scores with conventional measures of cognition ($r = .56$ with Mini-Mental Status Evaluation [MMSE; Folstein & Folstein, 1975]; $r = .59$ with Clock Drawing Test [CDT; Freedman et al., 1994]; and $r = .32$ with the visual attention Star Cancellation subtest of the Behavioral Inattention Test [BIT; Wilson, Cockburn, & Halligan, 1987]), and moderate correlations were found with caregiver ratings of ADLs ($r = .53$) and IADLs ($r = .58$). This initial study provided support for the validity of the instrument. Therefore, the purpose of this study was to further examine the reliability and validity of the Kettle Test in a geriatric stroke population faced with similar concerns regarding the impact of cognitive deficits on daily living abilities at discharge from rehabilitation.

The study was divided into two stages: (1) The first consisted of the study of interrater reliability of the Kettle Test between two sets of raters in two rehabilitation settings, and (2) the second consisted of the study of different aspects of validity in a sample of people at discharge from stroke rehabilitation and healthy elderly control participants. The objective of the first stage was to examine the
interrater reliability of the Kettle Test scores as administered by certified occupational therapists to stroke patients in rehabilitation. The objectives of the second stage were to examine

- The construct validity of the Kettle Test, in terms of its ability to differentiate between scores of people after stroke at discharge from rehabilitation and scores of age-matched healthy control participants who live independently in the community;
- The convergent validity of the Kettle Test, calculating relationships between Kettle Test scores with conventional measures of cognition; and
- The ecological validity of the Kettle Test, examining the relationship of safety level and BADL status at discharge and IADL status at home, 1 month after discharge

Method

Stage 1: Interrater Reliability Study

Participants were recruited from two centrally located geriatric rehabilitation hospitals in Jerusalem and Tel Aviv. Inclusion criteria required participants to be admitted to a rehabilitation hospital within 1 month after stroke, be ages 60 or older, be alert and without receptive aphasia, and live independently in the community before stroke. Twenty-one patients consecutively admitted to each rehabilitation hospital who met inclusion criteria participated in the reliability study (Hospital 1: n = 10; 4 men, 6 women; mean age = 79.3, standard deviation [SD] = 5.8. Hospital 2: n = 11; 6 men, 5 women; mean age = 77.82, SD = 5.1). The study was approved by the Human Rights Helsinki Committee of each hospital, and all participants signed informed consent. Four certified and experienced occupational therapists (two in each hospital) rated patient performance on the Kettle Test. The test was administered by one of the raters while the other observed the assessment. Independent ratings were then conducted immediately after the assessment. Interrater reliability (Spearman correlation coefficient) was found to be high for the Kettle Test total scores for both sets of raters (Hospital 1: r = .851, p = .001; Hospital 2: r = .916, p = .000).

Stage 2: Validity Study

Participants. Thirty-six stroke patients (18 with right hemisphere stroke and 18 with left hemisphere stroke) were recruited from a geriatric rehabilitation hospital before their discharge (mean length of stay = 63.1 days, SD = 29.2). Inclusion criteria were the same as delineated in the reliability study. In addition, 36 healthy control participants were recruited from a convenience sample of healthy elderly volun-

teers. The inclusion criteria for the control participants were age 60 or older, living independently in the community, and scoring >23 on the MMSE; see Table 1 for demographic data). The study was approved by the Human Rights Helsinki Committee of the hospital, and all participants signed informed consent.

Instruments. A short battery of conventional standardized measures of cognition recommended in geriatric stroke rehabilitation (Adunskey, Fleissig, Levenkrohn, Arad, & Noy, 2002) was used in this study. The battery included the following measures:

- The MMSE is a universal measure of cognitive status that assesses a broad range of basic cognitive abilities including orientation, attention, recall, working memory, spatial abilities, and language. Scores range from 0 to 30, and scores >23 are considered the cutoff for cognitive impairment.
- The CDT is widely used in cognitive screening for many neurological conditions (Freedman et al., 1994). In the stroke population, several scoring systems of a free-style drawing of a clock with hands at 10 past 11 have been shown to have construct validity and correlate with multiple cognitive domains, such as abstract thinking, executive functioning, and visuospatial construction (Suhr, Grace, Allen, Nadler, & McKenna, 1998). In this study, we used the scoring system of Rouleau, Salmon, Butters, Kennedy, and McGuire (1992), with a score range from 0 to 10, which was found to be reliable and valid in stroke rehabilitation (Suhr et al., 1998).
- The Star Cancellation subtest of the BIT was used to measure visual attention, as recommended for use in stroke rehabilitation practice (Edwards et al., 2006).
- The Cognitive scale of the FIM™ (CognFIM; Granger, 1998) was used as a measure of cognitive–functional status as expressed in daily activities (based on general observation) in five areas of cognition (expressive and receptive language, problem solving, social interaction, and memory). Each area is rated on a 7-point scale; total score ranges from 5 to 35. Reliability and validity have been extensively established in the stroke population (Heruti, Lusky, & Dankner, 2002).
- The functional outcome battery included the areas of BADLs, IADLs, and safety—outcomes necessary for independent community living.
- BADLs were measured with the Motor scale of the FIM (Granger, 1998) because it is a universal, reliable, and valid measure of rehabilitation outcome. The FIM Motor scale includes 13 items of basic self-care rated on a 7-point scale (total score range = 13–91).
- IADLs were measured with the IADL scale (Lawton & Brody, 1969; Lawton, Moss, Fulcomer, & Kleban, 1982),
Table 1. Demographic Variables and Performance on the Kettle Test by Group

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Stroke</th>
<th>Control</th>
<th>Results and Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>22 (61)</td>
<td>7 (19)</td>
<td>(\chi^2(1) = 12.99) (p = .000)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (39)</td>
<td>29 (81)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36 (100)</td>
<td>36 (100)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>Range: 60–89 M (SD): 74.81 (7.32)</td>
<td>Range: 60–84 M (SD): 72.67 (6.59)</td>
<td>(t(70) = 1.30, ns) (t(70) = -2.31, p = .024)</td>
</tr>
<tr>
<td>Education (years)</td>
<td>0–19: 10.44 (5.28)</td>
<td>0–19: 10.44 (5.28)</td>
<td></td>
</tr>
<tr>
<td>Kettle Test scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score(a)</td>
<td>Range: 1–29 M (SD): 9.34 (5.79)</td>
<td>Range: 0–3 M (SD): 0.42 (0.91)</td>
<td>(F(1,60) = 63.53, p = .000)</td>
</tr>
<tr>
<td>Assistance level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>5 (17)</td>
<td>35 (97)</td>
<td>(\chi^2(3) = 43.53) and 3.53 (t = .000)</td>
</tr>
<tr>
<td>Mild assistance</td>
<td>13 (45)</td>
<td>1 (3)</td>
<td></td>
</tr>
<tr>
<td>Considerable assistance</td>
<td>11 (38)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

Note. \(m = \) mean; \(SD = \) standard deviation; \(ns = \) not significant.

\(a\)Higher scores represent more assistance needed.

\(b\)Analysis of covariance, controlling for years of education.

a widely used measure for independence in extended areas of ADLs, including telephone use, shopping, food preparation, housekeeping, laundry, transportation, taking medication, and financial management (total score range = 0–23).

- Safety was measured with the Safety Rating scale, which is part of the Routine Task Inventory (RTI–E), a structured observation used to rate cognitive levels in daily routine activities (Allen et al., 1992; Karz, 2006). The 4-point scale ranges from 3 (unable to recognize the need for safety precautions) to 6 (anticipates hazards and plans safety procedures). The reliability and validity of the RTI–E has been studied in populations with cognitive disabilities (Karz, 2006).
- The Fugl-Meyer Motor Assessment (FMA; Fugl-Meyer, Jääsko, Leyman, Olsson, & Steglind, 1975; Rabadi & Rabadi, 2006) Upper-Extremity scale was used to measure motor status to evaluate the possible confounding relationship between motor function after stroke with performance on the Kettle Test.

Procedure

The assessment battery was administered within the last week before discharge from the rehabilitation hospital. The Kettle Test was administered by Hagit Harel. The other assessments were administered by experienced occupational therapists and rehabilitation professionals as part of the routine discharge assessment battery of the rehabilitation department. In addition, to examine IADL status 1 month after discharge, study group follow up included a telephone interview administered to caregivers. The follow-up assessment was completed on 29 participants, because 7 participants could not be reached. The Kettle Test scores of the 7 participants that were lost to follow-up were not significantly different from those of the other participants (\(t[34] = .892, p = .403\))

Data Analysis

Data were analyzed with SPSS (Version 15.0; SPSS, Inc., Chicago). Descriptive statistics and analysis of covariance (ANCOVA) were used to examine the group effect on the Kettle Test while controlling for years of education, which were found to differ significantly between groups. Pearson correlation analyses were used to examine the relationships among the cognitive and functional measures and motor and demographic variables. No significant differences were found between participants with right- and left-hemisphere stroke on the Kettle Test \(t[34] = 0.178, p = .860\); therefore, the analyses were conducted on the entire stroke group.

Results

Before studying validity, we examined the relationships of the Kettle Test with demographic and motor variables. The correlations between the Kettle Test scores and age, years of education, and FMA were all low and nonsignificant \((r_s = .04, -.14, \text{ and } -.19\), respectively). Regarding gender, no significant differences were found between men and women on the Kettle Test \(t[34] = 4.35, p = .67\).
### Construct Validity

As seen in Table 1, the Kettle Test scores of the stroke group were significantly higher than those of the control group, showing the need for more assistance on the test. ANCOVA, controlling for years of education (because a significant difference was found between groups on this variable), demonstrated a large significant group effect on the test ($F[1, 60]$ = 63.53, $p = .000$). The study group showed a wide range of performance on the Kettle Test (1–29), whereas the control group showed a narrow range (0–3), and only 1 control participant required mild assistance, demonstrating a floor effect on the test for this group.

### Convergent Validity

The mean scores of the study group on the conventional battery of cognitive measures are presented in Table 2. Results show a wide range of performance on all measures. The Kettle Test, which targets cognitive abilities in a functional context, was found to moderately significantly correlate with these conventional cognitive measures (Table 2). The correlations of the Kettle Test with the four cognitive measures ranged from .478 to .659 (all significant at $p < .01$), the highest being with the CognFIM, which is a measure based on observation in ADLs.

### Ecological Validity

The mean scores of the stroke group on the functional outcomes were as follows: 76.09 ($SD = 12.19$) on the FIM Motor scale, 4.94 ($SD = 0.75$) on the Safety Rating scale, and 10.70 ($SD = 5.07$) on the IADL scale at follow-up. The correlations of the Kettle Test and the other cognitive measures with these functional outcomes are presented in Table 3. The Kettle Test scores were significantly correlated with all three outcome measures, FIM Motor ($r = -.759$), Safety ($r = -.571$), and IADL 1 month after discharge at home ($r = -.505$). The correlations between the other cognitive measures and functional outcomes were lower, and no single measure correlated significantly with all three outcomes.

### Table 3. Pearson Correlation Coefficients of Functional Outcomes With the Kettle Test and Conventional Cognitive Measures

<table>
<thead>
<tr>
<th></th>
<th>BADLs (FIM Motor Scale)</th>
<th>Safety Level (RTI Safety Rating Scale)</th>
<th>IADLs (IADL Scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kettle Test</td>
<td>-0.759**</td>
<td>-0.571**</td>
<td>-0.505**</td>
</tr>
<tr>
<td>Mini-Mental Status Evaluation</td>
<td>0.261</td>
<td>0.122</td>
<td>0.401*</td>
</tr>
<tr>
<td>Clock Drawing Test</td>
<td>0.365*</td>
<td>0.112</td>
<td>0.182</td>
</tr>
<tr>
<td>Star Cancellation subtest of BIT</td>
<td>0.462**</td>
<td>0.102</td>
<td>0.237</td>
</tr>
<tr>
<td>FIM Cognitive scale</td>
<td>0.435**</td>
<td>0.446**</td>
<td>0.287</td>
</tr>
</tbody>
</table>

Note. BADLs = basic activities of daily living; RTI = Routine Task Inventory; IADLs = instrumental activities of daily living.

Discussion

The findings of this study provide initial support for the reliability and validity of the Kettle Test in stroke rehabilitation. We demonstrated that the Kettle Test—a structured observation of performance in a complex task—can be scored consistently among professional occupational therapists. In addition, participants after stroke required significantly more assistance than healthy control participants, who scored almost perfectly on the test. Moreover, the test performance of participants after stroke was not related to their motor status or educational background, but it was moderately correlated with their conventional cognitive test scores. Finally, Kettle Test scores were found to be significantly correlated with functional outcomes at the time of discharge from the rehabilitation hospital and at home.

The results concerning interrater reliability were encouraging, considering the difficulty of reliably assessing complex functional tasks. This positive finding can be attributed to the structured scoring scale of the observation (divided into discrete steps) and the expertise of the raters, who were occupational therapists trained in observations of functional performance. Further reliability analysis is in progress using videotape recordings that will enable comparisons among multiple raters. Another aspect of reliability pertaining to the stability of performance (test–retest reliability) was not deemed relevant for the test. Because the test incorporates an element of novel everyday problem solving, we hypothesized that immediate learning would occur from the exposure to the task and examiner cueing. Thus, similar to other tests that incorporate novel problem solving, performance is not expected to be identical in a subsequent immediate retest (Wilson, Alderman, Burgess, Emslie, & Evans, 1996).

The construct validity of the test was highly supported by the significant differences that were found between the stroke and control groups. This group effect could not be attributed to the difference in education because the effect
remained when controlling for years of education. Moreover, education was not found to be significantly related to performance on the Kettle Test, and no difference was found between the performance of men and women on the test. Therefore, the differences between the groups support the construct validity of the test to differentiate between known groups with identified differences in cognitive functional status. We expected that healthy adults who are living independently in the community would succeed on the Kettle Test; this expectation was indeed confirmed by the floor effect among the control group, in which only 1 participant needed mild assistance on the test. The findings suggest that the test can detect cognitive–functional deficits in adults with neurological dysfunction; however, the test is not sensitive to possible variance in cognitive–functional performance of normative healthy adults. Further research is required on larger, more representative samples of different diagnostic groups with identified cognitive disabilities.

The significant moderate correlations that were found with conventional established measures of cognition support the convergent validity of the Kettle Test, demonstrating the expected strength of association ($r = .50$) between measures with overlap of construct (Gregory, 2000). The correlations with the “table-top” tests (MMSE, CDT, and Star Cancellation) suggest that common underlying cognitive abilities, including attention, memory, praxis, and executive functions, are being tapped by these measures. However, the moderate degree of association supports the contention that the Kettle Test is also tapping into additional aspects of performance not addressed by these conventional cognitive measures.

These findings are similar to the initial findings regarding the Kettle Test in adults with suspected dementia (Hartman-Maeir et al., 2004) and to those of other studies of top-down assessments such as the AMPS and the EFPT, where moderate associations were found with table-top measures of cognition (Baum et al., 2008; Fisher, 2006a). The significant correlation that was found between the Kettle Test scores and the FIM supports the validity of the test as a measure of cognition in function. The FIM Cognitive scale provides a valuable measure of functional cognition in stroke rehabilitation; however, it relies on the day-to-day observations of clinicians in the rehabilitation setting and has a high focus on BADLs. This setting does not necessarily provide an opportunity to observe the client in the more complex tasks that are essential for independent living in the community after discharge. The moderate association that was found may shed light on the possible unique contribution of a short IADL performance test, beyond a general observation in daily life.

The ecological validity of the Kettle Test was substantially supported by the significant correlations with all functional outcomes (ADLs, safety, and IADLs). Conversely, the pattern of correlations between these outcomes and the conventional cognitive measures showed fewer and smaller effects, and no single measure correlated significantly with all outcomes. The outcome measures that were chosen for this study represent functional areas that are pertinent issues for discharge planning after stroke and for independent community living. The findings clearly support the advantage of the Kettle Test in its relation to these outcomes.

From our clinical experience, the ecological and face validity of the Kettle Test for discharge planning has been shown to be valuable for clients and their caregivers. On discharge from hospital, the need for assistance in ADLs and safety status is of major concern in planning for community living. Questions such as “Can I return home alone?” or “What type of assistance will I need to live independently in the community?” are of vital importance. For example, study participant A. R. is a 73-year-old married man who was living independently in the community before his stroke. He did not have any residual motor deficits at this stage of rehabilitation and was independent in BADLs, yet his conventional cognitive test scores revealed some mild deficits in measures of clock drawing and visual attention. A. R.’s performance on the Kettle Test demonstrated a significant need for assistance on multiple steps: He was baffled by the empty kettle, had difficulty connecting the electrical cord, and only prepared one cup of beverage (instead of two), using cold water. The face validity of this assessment for the client and his wife was useful in accepting the current need for assistance on going home.

Conclusions, Limitations, and Future Recommendations

The Kettle Test is a brief, cognitive–functional, top-down measure that assesses actual performance on a familiar IADL task with built-in complexity. The results of this study support the reliability and validity of this measure in stroke rehabilitation. The Kettle Test can be used in diverse settings; it is short, easy to learn and administer, and provides meaningful information pertaining to independent community living.

The study was limited by a relatively small sample size that did not enable multiple regression analysis for examining the unique prediction of the Kettle Test to the explained variance of functional outcomes beyond other conventional measures commonly used in geriatric stroke rehabilitation. In addition, the data regarding IADLs were collected by a telephone interview with caregivers, which may not have fully captured this area of functioning. Further studies should include home visits to confirm and replicate the current findings in the geriatric stroke population and in other diagnostic groups with cognitive disabilities.
References


**Essential Practice Guideline From AOTA Press!**

**Occupational Therapy Practice Guidelines for Adults With Stroke**

*By Joyce Sabari, PhD, OTR, FAOTA*

This evidence-based guideline details the significant contribution of occupational therapy in treating adults dealing with the limitations of having had a stroke. Appendixes include valuable resources such as *CPT™* codes related to occupational therapy for stroke survivors.

**Practice Guidelines can**

- Help clarify occupational therapy (OT) services for external audiences
- Help determine appropriateness of OT referrals
- Assist third-party payers in determining medical necessity for OT
- Assist OT educators in design of appropriate curricula.

ORDER #2211

AOTA Members: $49.00
Nonmembers: $69.50

ISBN: 978-1-56900-263-6

**Call 877-404-AOTA Shop Online at http://store.aota.org**