

Construct Validity, Criterion Validity, and Internal Consistency of GOT-Cog[®], a Novel Occupational Therapy Cognitive Screen

Emily Meise, John Corbett, Amanda Meyer, Pete Grevelding, Henry C. Hrdlicka

Importance: No single cognitive screen adequately captures the cognitive domains needed for inpatient occupational therapy treatment planning.

Objective: To assess the construct validity of the Gaylord Occupational Therapy Cognitive (GOT-Cog[®]) screen, a novel comprehensive cognitive screen that evaluates functional cognition.

Design: Randomized crossover controlled study design using the St. Louis University Mental Status (SLUMS) exam as a comparator.

Setting: Long-term acute-care hospital.

Participants: Participants were inpatients admitted to Gaylord Hospital who were ages 18 yr or older, prescribed occupational therapy services, with no documented history of dementia, Alzheimer's, or preexisting intellectual disability and no present aphasia.

Intervention: During participants' initial occupational therapy evaluation, either the SLUMS or GOT-Cog were randomly delivered; the screen that was not delivered on admission was delivered 22 to 26 hr later by the same or a different clinician.

Outcomes and Measures: GOT-Cog and SLUMS total scores and individual item and domain scores.

Results: Ninety-eight participants yielded sufficient data for analysis. Total GOT-Cog and SLUMS scores positively correlated ($p < .0001$). All shared domains between GOT-Cog and SLUMS were significantly correlated ($p \leq .0155$); similarly, all unique domains showed significant correlations with both GOT-Cog and SLUMS total scores ($p \leq .0194$). No ordering effects were observed ($p \geq .8081$). Despite having 11 more items, GOT-Cog took only 6 min longer to complete (10 vs. 16 min; $p < .0001$). Both demonstrated adequate internal consistency.

Conclusions and Relevance: The GOT-Cog has overall strong construct and criterion validity. Going forward, we will evaluate the rater reliability and responsiveness of the GOT-Cog.

Plain-Language Summary: Occupational therapists evaluate clients' cognitive strengths and limitations in relation to activities of daily living and instrumental activities of daily living. Occupational therapists use this evaluation to help clients identify strategies to adapt to their specific environments, support their independence, and improve their ability to perform tasks. No cognitive screen currently exists that adequately evaluates a person's cognitive domains as part of treatment planning for inpatient occupational therapy. This study reviewed the construct validity, criterion validity, and internal consistency of the Gaylord Occupational Therapy Cognitive screen (GOT-Cog), a new comprehensive cognitive screen. The GOT-Cog was used with inpatients at a long-term acute-care hospital as part of their initial occupational therapy evaluation. The study found that the GOT-Cog has overall strong construct and criterion validity. Future studies will evaluate the interrater and intrarater reliability and responsiveness of this new cognitive screen.

Meise, E., Corbett, J., Meyer, A., Grevelding, P., & Hrdlicka, H. C. (2025). Construct validity, criterion validity, and internal consistency of GOT-Cog[®], a novel occupational therapy cognitive screen. *American Journal of Occupational Therapy*, 79, 7901205010. <https://doi.org/10.5014/ajot.2025.050832>

Occupational therapists evaluate cognition in relation to functional tasks (i.e., functional cognition) by utilizing occupation-based measures and completing

performance-based tests. This is conducted by evaluating clients' cognitive strengths and limitations in the context of activities of daily living (ADLs) and

instrumental activities of daily living (IADLs), allowing an occupational therapist to help clients identify strategies to compensate and adapt to their specific environments while improving their independence with ADLs and IADLs. Assessing cognition in the context of ADLs, IADLs, and other function-related tasks is necessary to identify cognitive impairments that may challenge a patient's ability to accomplish real-world tasks (American Occupational Therapy Association [AOTA], 2021). *Functional cognition* is the interaction of cognitive, self-care, and community-living skills. Specifically, it refers to the interaction between the thinking and processing skills needed to complete ADLs and IADLs, such as eating, bathing, dressing, household and financial management, medication management, volunteer activities, driving, and work (AOTA, 2013; Faul et al., 2010; Okie, 2005; Wolf et al., 2019). In this way, occupational therapists use everyday activities in familiar contexts, because they have a high potential for client engagement.

The role of occupational therapists in evaluating cognition has been a recent area of focus by the AOTA. The Improving Medicare Post-Acute Care Transformation Act (IMPACT Act) was signed into law in 2014 and requires skilled nursing facilities, inpatient rehabilitation facilities, long-term care hospitals, and home health agencies to report standardized patient assessment data with regard to quality measures. This includes the standardization of the Centers for Medicare and Medicaid Services (CMS) data fields for baseline, discharge, and change in physical and cognitive function for all patients (AOTA, 2014). In response to the IMPACT Act, the AOTA has advocated for the importance of functional cognition and its benefits in occupational therapy treatment planning.

Functional cognition is fundamental to achieve and maintain community placement, discharge stability, and prevent failed care transitions (AOTA, 2013; Giles et al., 2020). Given the value placed on evaluating and treating impaired functional cognition to prevent poor post-discharge outcomes, the occupational therapy profession must develop, validate, and systematically use performance-based assessments of functional cognition that include ADL and IADL elements (Wolf et al., 2019). The recognition of cognition in policy and our understanding of its role in successful performance and participation requires us to emphasize functional cognition in our daily practice, education, and research.

In the long-term acute care hospital (LTACH) setting, many patients are admitted with preexisting comorbidities or after a prolonged hospitalization. Assessing cognitive function is imperative for occupational therapy goal setting, as well as for determining a safe discharge plan. Identifying potential cognitive impairments early in a patient's admission can assist in timely referrals to speech therapy and/or neuropsychology, as well as assist care management in determining an appropriate discharge location.

Existing cognitive outcome measures are frequently designed to assess a specific diagnosis in a specific population. Although these measures are useful for these key populations, they often lack the ability to identify multiple areas of concern in a reasonable time for administration. Furthermore, many of the existing cognitive outcome measures available to occupational therapists were created to assess dementia or mild neurocognitive impairment in the community-dwelling population, such as the Montreal Cognitive Assessment (MoCA) and the St. Louis University Mental Status exam (SLUMS; Nasreddine et al., 2005; Tariq et al., 2006).

Previously, we reported on the development and content validity of a novel cognitive screen we have named the Gaylord Occupational Therapy Cognitive screen (GOT-Cog[®]; Hrdlicka et al., 2024). This was, in part, in response to the costly certification process of the MoCA. Although our inpatient occupational therapy department had made the choice to switch to the SLUMS, our staff found that it did not meet their needs. The SLUMS and the MoCA primarily assess memory, but they do not address cognition in the context of functional tasks or patient's daily activities. This limits their overall utility for developing an appropriate inpatient occupational therapy treatment plan on the basis of clients' functional cognition, thus establishing the need for the development of a new cognitive screening tool (Hrdlicka et al., 2024). This new comprehensive cognitive screening measure integrates many of the essential cognitive domains necessary for occupational therapy treatment and discharge planning, such as orientation, verbal fluency, visuospatial tasks, functional problem-solving, reasoning, sequencing, attention, divided attention, and immediate and long-term memory recall (Hrdlicka et al., 2024). It does this in the context of functional tasks, relevant to patients' environment and daily activities. Using two iterative rounds of Delphi-style expert panel review, GOT-Cog was found to have overall excellent content validity (Hrdlicka et al., 2024). Given these findings, we proceeded to test the construct validity, criterion validity, and internal consistency of the GOT-Cog.

Construct validity assesses the degree to which a scale measures the idea or concept it claims to evaluate (i.e., functional cognition). *Criterion validity* assesses the degree to which the scores on a measurement tool, such as a cognitive screen, are related to a specific outcome that it was designed to measure. *Internal consistency reliability* evaluates the extent to which individual items of a multi-item scale succeed at measuring different aspects of the same broader construct. These forms of testing are conducted by comparing the scale in question with another validated scale with similar qualities and calculating how well the scales correlate. The aim of this study is to expand on our previous work and evaluate the construct validity, criterion

validity, and internal consistency reliability of GOT-Cog. Last, we establish cut-off scores for GOT-Cog.

Method

Ethics Review

Before data collection, the study was submitted and reviewed by the Gaylord Hospital Institutional Review Board (IRB). Given the low-risk nature of the study, it was given an exempted research status and was approved by the Gaylord Hospital IRB in October 2021.

Study Design

To evaluate the construct validity, criterion validity, and internal consistency reliability of GOT-Cog, a randomized cross-over controlled study design was developed and conducted at Gaylord Specialty Healthcare, an LTACH located in Wallingford, CT. Criterion and construct validity were assessed using the SLUMS as a comparator (Tariq et al., 2006). The SLUMS is a standardized measure of cognition whose reliability and validity have been well documented (Noyes et al., 2023; Tariq et al., 2006). The SLUMS was selected as the reference standard, because it shares many of the same qualities and domains as the GOT-Cog. Moreover, the SLUMS was already implemented at the study site, and its use reduced the burden on the clinical investigators conducting the study and collecting data.

At the start of the study and whenever additional occupational therapists volunteered to assist with data collection, occupational therapists were provided with an in-person, in-service training on how to administer the SLUMS; in addition, they were required to watch the Veterans Affairs-produced video available online, as is recommended by Saint Louis University (n.d.). All occupational therapists also attended an in-service training on how to administer the GOT-Cog. At this stage of development, the rater reliability of the GOT-Cog was not evaluated and will be addressed in future works.

Once recruited, participants were assigned to a primary occupational therapist involved in the study, and they were randomly delivered either the SLUMS or the GOT-Cog during their initial occupational therapy assessment. On the following day, within 22 to 26 hr, the second scale, which was not delivered the day before, was administered by either the same or another occupational therapist involved in the study. The 22- to 26-hr range was chosen to minimize the mental fatigue of a same-day administration of the second scale. This allowed leeway to complete the follow-up assessment on a consecutive day, at a similar time and location, while accounting for practical issues such as scheduling difficulties and meal times.

Participants

Participants were recruited from all inpatients admitted to Gaylord Hospital who were ages 18 yr and

older and who were prescribed occupational therapy services. To be considered an eligible participant, patients had to meet all study criteria outlined in Figure 1. Because of the infection prevention guidelines implemented during the COVID-19 pandemic, we were unable to enroll patients with an active COVID-19 infection into the study. This decision was made to prevent potential contamination, because bringing testing forms in and out of the room was deemed unfeasible.

In addition to these criteria, as determined by the clinical judgment of the occupational therapist involved in the study, anyone who presented with an observable and distinguishable change in demeanor between assessment periods was withdrawn from the study, and a new candidate was recruited in their place. This included, but was not limited to, a significant change in confusion or mental status, a significant change in laboratory values, significant sleep disturbances, new O₂ requirements that would affect performance, and medication interactions affecting baseline mental status.

Outcome Measures and Data Collection

The primary outcome measures were the SLUMS and the GOT-Cog. Total scores, individual item scores, and domain scores were all collected and recorded separately for analysis.

Statistical Analyses

Data were analyzed using GraphPad Prism, Version 10.0.0 (GraphPad Software) and R Studio, Version 4.2.2 (Posit PBC). Descriptive statistics and 95% confidence intervals (95% CIs) were used to describe population characteristics. All data sets were analyzed using normality and lognormality testing to evaluate if there were violations in the equal variances assumptions between groups. Nonparametric testing was used as necessary after the results of normality and lognormality testing. Statistical significance was set at an α level of .05 for all statistical tests.

Ordering Effects, Completion Time, Range Utilization, and Education Effects

We used criterion and concurrent validity testing to show how well the SLUMS and GOT-Cog correlate with one another. This was measured using correlation coefficients and least squares regression analysis to evaluate whether participants achieved the same or similar outcomes between scales. Because of nonnormal distribution, as shown by using the Shapiro-Wilks test, order effects were determined by using the Mann-Whitney *U* test. Completion time between the two screens was evaluated using a Wilcoxon matched-pairs signed-rank test. Floor and ceiling effects were determined by calculating the percentage of participants who scored the maximum (ceiling effects) and the percentage of participants who scored the minimum (floor effects).

Figure 1. Study Criteria

Inclusion Criteria
<ul style="list-style-type: none"> • Be an adult age 18 yr or older • Be an admitted inpatient to the study site with an order for occupational therapy • Ability to be alert, be oriented, and attend to cognitive testing verbally and in written format • Ability to understand and respond to simple verbal instructions and one-step commands in English well enough to participate in the occupational therapy evaluation without interpreter services • Able to complete two assessments on 2 consecutive business days
Exclusion Criteria
<ul style="list-style-type: none"> • Be a member of a vulnerable population, including inmates, women who are known to be pregnant, and children under age 18 yr • Aphasia limiting the comprehension or communication of the cognitive assessments • A previously documented diagnosis of dementia, Alzheimer's disease, or preexisting intellectual disability • Any reason why the physician may deem it harmful that the participant enrolls or continues in the study • Inability to complete the two cognitive assessments on consecutive days • Motor function impairment that limits motor tasks of the examination, even if modified • Inability to read or complete exam because of preexisting visual impairment
Stopping Criteria
<ul style="list-style-type: none"> • Significant changes in level of confusion or mental status between assessments • Significant changes in laboratory values • Patient study participation put on hold per medical team • Significant sleep disturbances • New O₂ requirements that would affect performance • Major medication interactions

Before running a correlation analysis against the assessment total score for education effects, we coded individual self-responses of having the equivalent of a high school diploma or greater as 1, and responses of having less than the equivalent of a high school diploma were coded as 0. These were then correlated against the respondents' final scores using point-biserial correlation (PBC). We conducted an analysis of the education subgroup (i.e., all education, high school diploma or higher, or less than a high school diploma) using Pearson's correlation (*r*) analysis. We used an ordinary one-way analysis of variance (ANOVA) to determine whether there were any within-group differences in mean total scores by all education subgroups.

Convergent and Divergent Validity

Before analysis, domains were determined to be either shared between tools or unique to an individual tool. This was determined on the basis of domain structure (i.e., wording), intent, and presentation of each item to recipients. For the SLUMS, we referenced the resources provided by St. Louis University, which were then compared with GOT-Cog (Saint Louis University, n.d.). Shared domains were then evaluated for

convergent validity and unique domains were evaluated for divergent validity.

To calculate the convergent and divergent validity, we implemented a multitrait-multimethod matrix analysis. This method compares the scores of domains and items measured by different tests and shows how well items and domains expected to overlap between assessments correlate. Conversely, those domains and items that were not expected to overlap should have a low or negative correlation. It is for this testing that we collected the individual responses to each question.

We assessed the convergent validity of shared domains using least squares regression and Pearson's correlation analysis. Differences in shared domains that were scored on the same scale were analyzed using Mann-Whitney *U* testing. To assess the divergent validity of unique domains, we compared the scores for each domain and correlated them to the respective total SLUMS and GOT-Cog scores using Spearman's correlation analysis. A hypothetical slope = 0, indicating no correlation, was used to test the fit of regressions and a Fisher's *r*-to-*z* transformation test was used to compare the correlation coefficients.

Internal Consistency

First, item difficulty (i.e., percent correct) was determined by dividing the item average score by the total points possible for that item. We used PBC to determine the strength and relationship of each item to the total score of its respective measure. For this analysis, all items with a maximum score greater than 1 point were recoded so that responses with the maximum score were recoded as 1 and all partial or incorrect responses were coded as 0. For items that were not scored out of a maximum of 1 point, item-to-total score correlations (ITTCs) were also conducted. The r values of .20 or higher were considered to display adequate internal consistency. Last, Cronbach's α coefficient was calculated for each assessment. In addition, we calculated the α coefficient when each item was sequentially deleted. If the alpha coefficient if the item is deleted (AIID) is larger than the scale-level α coefficient, this indicated that the item may be redundant and may possibly introduce additional variation or error to the scale in total. Taken together, items with PBC and ITTC r values less than .20 and AIIDs higher than the scale-level α coefficient may be evaluated for revision or elimination.

Distribution-Based Cutoff Score

Determining appropriate cutoff scores for the GOT-Cog screen was important for treatment planning and clinical decision-making. Because the GOT-Cog and SLUMS are on different scales, sensitivity and specificity testing would be inappropriate; therefore, we used a bootstrap distribution-based cutoff analysis. All GOT-Cog scores were randomly sampled with replacement 100 times to get a random sample of scores, and this process was repeated 10,000 times to develop a bootstrap distribution. This process allowed us to enhance and identify underlying patterns in the data that were not initially recognizable in the original dataset because of sampling error. A bootstrap distribution was developed for the 25th quartile, 50th quartile (median), and 75th quartile. The center of these bootstrap distributions for each quartile were then the basis for the cutoff scores for the GOT-Cog moving forward, with clinical judgment used to determine which side of the distribution to designate the cutoff criteria for the determination of minimal or no, mild, moderate, and severe deficits.

Results

Participant Characteristics

This study included participants who were admitted to the inpatient rehabilitation program with an order for occupational therapy. They were recruited from December 2021 through April 2022. In total, 129 eligible patients were approached for data collection, with 102 participants completing at least one assessment, and 98 participants completing both assessments (see

Figure A.1 in the Supplemental Material, available online with this article at <https://research.aota.org/ajot>).

Of these 98 participants, 44.9% (44/98) were female, the mean age of the cohort was 59.8 yr (95% confidence interval [CI] [56.5, 63.2]), the median age was 63 yr (range = 18–90), the mean length of stay was 22.7 days (95% CI [20.4, 25.0]), and 90.8% (89/98) had received a high school diploma or the equivalent. These patients represented 10 clinical programs, with the stroke (22.4%; 22/98) and young stroke (17.3%; 17/98) populations being the most represented (Table 1).

Ordering Effects, Completion Time, Range Utilization, and Education Effects

To first rule out ordering effects, the means of each scale were compared with those delivered first or second. On average, there were no differences between the mean scores and 95% CI of the GOT-Cog screens delivered on Day 1 versus Day 2 ($M = 24.9$, 95% CI [23.2, 26.6] vs. $M = 24.7$, 95% CI [22.7, 26.8], $p = .8691$), as well as mean scores of the SLUMS delivered on Day 1 vs. Day 2 ($M = 20.5$, 95% CI [18.9, 22.1] vs. $M = 20.9$, 95% CI [19.4, 22.5], $p = .8081$). This indicated that test order did not affect the average score for either the GOT-Cog or the SLUMS.

Table 1. Patient Characteristics

Characteristic	Value
Age, yr, M [95% CI]	59.8 [56.5, 63.2]
LOS, days, M [95% CI]	22.7 [20.4, 25.0]
SLUMS score, M [95% CI]	20.7 [19.6, 21.8]
GOT-Cog score, M [95% CI]	24.8 [23.5, 26.1]
Sex, n (%)	
Female	44 (44.9)
Male	55 (55.1)
Education, n (%)	
HS diploma equivalent or higher	89 (90.8)
Less than HS diploma equivalent	9 (9.2)
Program, n (%)	
Brain injury	11 (11.2)
Stroke	22 (22.4)
Young stroke	17 (17.3)
Spinal cord injury	5 (5.1)
Orthopedic	8 (8.2)
Pulmonary rehabilitation	5 (5.1)
Neurorehabilitation	4 (4.1)
Wound, medical	4 (4.1)
Pulmonary, medical	9 (9.2)
Other general complex, medical	13 (13.3)

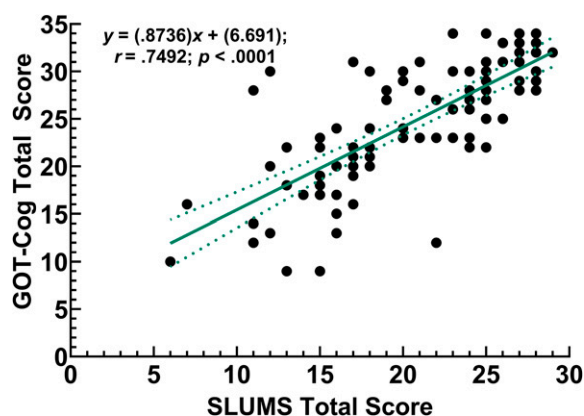
Note. CI = confidence interval; GOT-Cog = Gaylord Occupational Therapy Cognitive screen; HS = high school; LOS = length of stay; SLUMS = St. Louis University Mental Status.

The completion time for each scale was also evaluated. Despite the GOT-Cog being 11 items longer, it took only 6 min longer ($p < .0001$) to complete, on average, than the SLUMS ($M = 16.4$ min, 95% CI [15.7, 17.1] vs. $M = 10.5$ min, 95% CI [9.4, 11.6]). Of the 98 participants who completed both assessments, 76.7% (23/30) of the SLUMS scoring scale and 76.5% (26/34) of the GOT-Cog scoring scale were used. Although 4.1% (4/98) of patients who completed the GOT-Cog reached the assessment ceiling, this is still below the accepted cutoff of 15%; in comparison, no patients reached the SLUMS ceiling (see Table A.1 in the Supplemental Material; Portney & Gross, 2020).

Because the SLUMS scoring schema is dependent on patient education, the influence of education on total assessment scores was also evaluated (Tariq et al., 2006). No significant correlation between education level and total score was observed for the GOT-Cog ($r_{pb} = .077$, 95% CI [-.123, .272], $p = .4469$). Also, notably, no significant correlation between education level and total score was observed for the SLUMS ($r_{pb} = .067$, 95% CI [-.133, .262], $p = .5111$; Table A.2).

Furthermore, the correlation of GOT-Cog and SLUMS scores by education subgroup (i.e., all education, high school diploma equivalent or higher, or less than high school diploma) was also conducted. GOT-Cog and SLUMS total scores, regardless of education level, were significantly positively correlated ($r_{pb} = .749$, 95% CI [.647, .825], $p < .0001$; Figure 2; Table A.3). The GOT-Cog and SLUMS total scores of individuals with less than a high school diploma or the equivalent were also significantly positively correlated ($r_{pb} = .832$, 95% CI [.376, .964], $p = .0054$), as well as individuals with a high school diploma or higher ($r_{pb} = .745$, 95% CI [.635, .825], $p < .0001$; Table A.3).

Figure 2. Correlation of SLUMS and GOT-Cog Total Scores



Note. For the 98 participants who completed both the SLUMS and GOT-Cog, the total scores were plotted, and a linear regression was plotted using straight-line least squares fit. Solid line represents the calculated regression line; dashed lines represent the 95% confidence interval bands. GOT-Cog = Gaylord Occupational Therapy Cognitive screen; SLUMS = St. Louis University Mental Status.

Within-group analysis by one-way ANOVA further indicated that SLUMS ($p = .8042$) and GOT-Cog ($p = .7468$) scores were not affected by education level in this cohort. Given this, all further analyses included the total dataset and do not distinguish between patients with or without a high school diploma or the equivalent.

With the data taken together, the significant strong positive correlation observed between SLUMS and GOT-Cog total scores indicated the convergent validity of GOT-Cog.

Convergent Validity of Shared Domains

Between the GOT-Cog and SLUMS, seven domains share similarly structured items: Orientation, Delayed Recall, Verbal Fluency, Visuospatial, Attention, Auditory Memory, and Problem-Solving. To further assess the convergent validity of the GOT-Cog and SLUMS, we evaluated the correlation of the subscore for each of these shared domains (Table 2). All shared domains displayed a significant positive correlation ($p \leq .0155$), further demonstrating convergent validity.

Serendipitously, the GOT-Cog Short-Term Delayed Recall and Auditory Memory domains were scored using the same scale as that used for the corresponding SLUMS domain (5 points and 8 points, respectively). Furthermore, a subcomponent of the GOT-Cog, Attention domain Item 4, was also scored the same as the corresponding SLUMS domain (2 points). These similarities allowed us to directly compare the scores of these shared and similarly scored domains between assessments (Table A.4).

When compared, the average scores for both the shared Auditory Memory domain and Attention subdomain were not significantly different between the SLUMS and GOT-Cog ($ps = .2533$ and $.1723$, respectively). In contrast and of note, the GOT-Cog Short-Term Delayed Recall domain scores were significantly higher than those of the SLUMS ($M = 3.4$, 95% CI [3.0, 3.81] vs. $M = 2.6$, 95% CI [2.3, 2.9], $p = .0001$). This is of note, because the GOT-Cog takes, on average, 6 min longer to complete and requires a set 10 min to lapse before assessing recall.

Divergent Validity of Unique Domains

Not all domains were shared between the SLUMS and the GOT-Cog. Unique to the SLUMS was the Executive Function domain, and unique to the GOT-Cog were the Divided Attention and Sequencing domains. To assess the divergent validity of these unique domains, we compared the scores for each domain and correlated them with the respective SLUMS and GOT-Cog total scores. All unique domains showed significant positive correlations with the total scores of both assessments ($p \leq .0194$), with each unique domain having a larger r value when correlated with the assessment for which they were originally designed. When the r values were compared, the only

Table 2. Correlations and Agreement of Shared Domains

Domain, Item Number, and Points Possible		Pearson <i>r</i> [95% CI], <i>p</i>
SLUMS	GOT-Cog	
Attention, Immediate Recall, and Orientation; Items 1–3; 3 points	Orientation; Items 5–9; 5 points	.244 [.048, .422], .0155
Delayed Recall with Interference; Items 7 and 4; 5 points	Short-Term Delayed Recall; Items 22 and 10; 5 points	.537 [.379, .665], <.0001
Memory/Recall (with time constraint); Item 6; 3 points	Verbal Fluency; Items 1 and 2; 2 points	.692 [.572, .783], <.0001
Visuospatial; Question 9; 4 points	Visuospatial; Items 11 and 12; 2 points	.489 [.321, .624], <.0001
Registration and Digit Span; Item 8; 2 points	Attention; Items 3 and 4; 4 points	.361 [.175, .522], .0003
Executive Function Plus Extrapolation; Item 11; 8 points	Auditory Memory/Attention; Items 14–17; 8 points	.389 [.207, .545], <.0001
Numeric Calculation and Registration; Item 5; 3 points	Functional Problem Solving; Items 19–21; 6 points	.448 [.274, .594], <.0001

Note. CI = confidence interval; GOT-Cog = Gaylord Occupational Therapy Cognitive screen; SLUMS = St. Louis University Mental Status exam.
^a*n* = 98.

significant difference between correlations was for SLUMS Item 10 and its correlation with the SLUMS and GOT-Cog total scores ($z = 1.96$, $p = .05$). Although GOT-Cog Items 13 and 18 have marginally greater r values when correlated with GOT-Cog total scores versus SLUMS total scores, these r values were not significantly different when correlated with SLUMS scores (Item 13, $z = -0.69$, $p = .4902$; Item 18, $z = -0.25$; $p = .8026$; Table A.5).

Internal Consistency: Item Difficulty, Point-Biserial Correlation, Item-to-Total Correlation, and Cronbach's α if Item Is Deleted

Internal consistency, a form of reliability testing, was also completed for the SLUMS and GOT-Cog. The SLUMS was also included in this analysis to demonstrate the item discrimination of the SLUMS in the LTACH population compared with that of the GOT-Cog. First, item difficulty and PBCs were calculated for each item. If the item total score was greater than 1 point, the ITTC was also calculated. For the SLUMS, Item 3 (i.e., “What state are we in?”), representing the Orientation domain, had a percent correct rate of 98.1% and a PBC Pearson's r of .181 (95% CI [−.013, .362], $p = .0666$; Table A.6). All other PBCs and ITTCs for the SLUMS displayed a significant positive correlation, $r \geq .268$, and a percent correct rate ranging from 51.8% to 96.1%.

Similarly, GOT-Cog Item 7 (i.e., “What is the year?”), representing the Orientation domain, had a percent correct rate of 97.0% and a PBC Pearson's r of .166 (95% CI [−0.031, 0.351], $p = .0980$; Table A.7). All other PBCs and ITTCs for the GOT-Cog displayed a significant positive $r \geq .236$ and a percent correct rate ranging from 46.0% to 97.0%.

AIID and the change in AIID (Δ AIID) compared with the intact scale were also calculated. These can be

used to determine whether an item may be redundant and possibly introduce variation and error to the scale in total. For SLUMS Items 3 and 11 (i.e., “I am going to tell you a story ...”)—representing the SLUMS Extrapolation and Executive Function domains, respectively— Δ AIID = −.005 and −.011, respectively; for all other SLUMS items, Δ AIID ≥ 0.001 (Table A.8). Similarly, for GOT-Cog Item 7, Δ AIID = −.001; for all other GOT-Cog items, Δ AIID $\geq .0001$ (Table A.9).

GOT-Cog Cutoff Criteria: Distribution-Based Modeling

To provide direction to use the GOT-Cog for treatment planning, it was necessary to establish cutoffs. Because the GOT-Cog and SLUMS are scored on two different scales (34 and 30 points, respectively), sensitivity and specificity testing to determine cutoffs would be inappropriate. As an alternative, we used a bootstrap distribution-based method to determine cutoffs. First, the population of GOT-Cog total scores for the 98 complete datasets was determined (Figure A.2). These 98 complete datasets were then randomly sampled with replacement 100 times to generate a single random sample from the population. This process was then repeated 10,000 times, and each quartile was plotted to generate a bootstrap distribution of the quartiles. In calculating the center of each bootstrap distribution of quartiles, we found that the 25th quartile = 21, the 50th quartile = 26, and the 75th quartile = 30.

GOT-Cog Cutoff Criteria: Recommendations

On the basis of these findings and clinical judgment, we recommend that patients with GOT-Cog scores ranging from 0 to 21 should be considered to have

severe deficits in functional cognition; those with scores ranging from 22 to 25 should be considered to have moderate deficits in functional cognition; those with scores ranging from 26 to 29 should be considered to have mild deficits in functional cognition; and those with scores ranging from 30 to 34 should be considered to have minimal or no deficits in functional cognition.

Discussion

Principle Findings and Interpretation

Screening functional cognition has been a recent area of focus by the AOTA. In response to the IMPACT Act, the AOTA has advocated for the importance of functional cognition and its benefit in occupational therapy treatment planning (AOTA, 2014, 2021). The IMPACT Act requires skilled nursing facilities, home health agencies, inpatient rehabilitation facilities, and long-term acute-care hospitals to report standardized patient assessment data with regard to quality measures, resource use, and other measures (AOTA, 2014, 2021). The CMS specifies that functional status, cognitive function, changes in cognitive function, and discharge to community should be standardized under the IMPACT Act, in addition to other domains (AOTA, 2014, 2021). AOTA shared with CMS that, under the umbrella of cognition, assessment of functional cognition is necessary to identify cognitive impairments that challenge the patient's ability to accomplish real work tasks (AOTA, 2014, 2021). Occupational therapists are experts in measuring functional cognition, which includes assessing everyday task performance, and occupational therapists treat cognitive impairments, because these impairments have the potential to compromise patients' safety and long-term well-being (AOTA, 2014, 2021).

The GOT-Cog is a comprehensive cognitive screen that assesses patient cognition in a context related to a patient's daily activities with an increased focus on functional problem-solving and sequencing compared with previous screens, such as the MoCA and SLUMS. The MoCA and SLUMS were created to assess dementia or mild neurocognitive impairment in community-dwelling populations, with a heavy focus on memory and attention. However, neither assesses cognition and memory in relation to functional tasks, which is one way in which the GOT-Cog is distinct from other tools. The GOT-Cog focuses on screening cognition through functional tasks and in the context of patient's daily activities; for example, screening immediate and short-term delayed recall with a 10-min timed delay using a grocery list rather than random words, as well as assessing functional problem-solving through medication and money management, which allows the patient to solve the problem by using pen and paper and not only by using working memory like on the SLUMS Item 5, where the patient is asked to solve a grocery store task and complete computation

with no pen and paper. Furthermore, the GOT-Cog was developed with the aim of assessing patients who are seen at the LTACH level of care with dynamic diagnoses such as stroke, brain injury, spinal cord injury, and other medically complex diagnoses. Although the GOT-Cog takes slightly more time to complete, it is more a comprehensive cognitive screening tool related to functional tasks.

Previously, the GOT-Cog was found to have overall excellent content validity and the goal of the present study was to determine the construct validity, criterion validity, and internal consistency of the GOT-Cog. To establish criterion and construct validity, we compared the total scores and individual domain scores of the GOT-Cog and SLUMS.

Overall, GOT-Cog total scores were shown to correlate positively with SLUMS scores. Between the GOT-Cog and the SLUMS, seven domains were found to have similarly structured items: Orientation, Delayed Recall, Verbal Fluency, Visuospatial, Attention, Auditory Memory, and Problem-Solving. All shared domains displayed a significant positive correlation. Not all domains were shared between the SLUMS and the GOT-Cog. Although the GOT-Cog addresses executive functioning skills in the context of daily activities, the specific domain of Executive Function is unique to the SLUMS. Unique to the GOT-Cog are the Divided Attention and Sequencing domains. To assess the validity of these unique domains, the scores for each unique domain were correlated with their respective total SLUMS and GOT-Cog scores. This analysis found that all unique domains showed significant positive correlations with the total scores of both assessments. However, only the SLUMS's unique domain of Executive Function correlated more strongly with its native measure, whereas the GOT-Cog domains of Divided Attention and Sequencing demonstrated similar correlations with their native measure and the SLUMS. Together, this indicates that these unique domains measure the same overall construct (i.e., cognition). The SLUMS domain of Executive Function then demonstrates divergent validity between the SLUMS and GOT-Cog, whereas the unique GOT-Cog domains demonstrate further convergent validity between assessments. Together, these analyses indicate that the GOT-Cog has overall strong criterion and construct validity.

When evaluating internal consistency, we made the decision to evaluate and report the internal consistency of SLUMS, as well as that of GOT-Cog, because it had not been previously reported in the literature. For the SLUMS, the PBC-ITTC and AIIDA analysis indicated that SLUMS Item 3 (i.e., "What state are we in?") may be redundant in the context of the total scale and may possibly introduce variation and error into the total score. Although it is not redundant, AIIDA analysis indicated that SLUMS Item 11 (i.e., "I am going to tell you a story ...") could introduce variation and error into the total SLUMS score as well. The same PBC/

ITTC and Δ AIID analyses indicated that GOT-Cog Item 7 (i.e., “What is the year?”) may also be redundant in the context of the total scale and may possibly introduce variation and error into the total GOT-Cog score. Together, the internal consistency of the GOT-Cog was found to be sufficient, and Item 7, the only item found to be potentially at odds, was maintained, because it was deemed to be clinically important.

It was not surprising that, being 11 items longer, the GOT-Cog took 6 min longer to administer than the SLUMS. However, despite this length, participants demonstrated a greater short-term delayed recall score in the GOT-Cog than in the SLUMS. This difference is, in part, likely due to the difference in how the domain is delivered in the GOT-Cog versus the SLUMS. In the SLUMS, five random objects are read aloud to the participant, and they are asked to remember them. They are then asked to recall those items three questions later. In the GOT-Cog, a list of five objects described as a grocery list is read aloud to the participant; the participant is then allowed to review the list for 1 min and is then asked to repeat the list once it is removed. Then, after at least 10 minutes have elapsed, the participant is asked to recall those five items. Delivering the cues verbally, allowing the participant to visually review the cues, and ensuring that a minimum of 10 minutes have elapsed may allow the patient to compensate for any underlying audiovisual deficits, thus providing a more accurate measurement of true short-term delayed recall.

Limitations

Several limitations of this work need to be addressed. First, because the study was completed in the LTACH setting, the GOT-Cog has only been validated for the LTACH population; therefore, it may not be valid to use in acute-care hospitals, inpatient rehabilitation facilities, skilled nursing facilities, and outpatient settings, and additional work is needed.

Second, the SLUMS was primarily designed for the people in the community-dwelling population who are facing dementia diagnoses. Although the SLUMS is widely used in the inpatient setting, a comparison of the GOT-Cog with the SLUMS is not an exact comparison, despite their similarities.

Third, the timing of the administration of the second screen was 22 to 26 hr after the initial assessment; as a result, several datasets were lost because of inaccurate timing between the two measures to meet study criteria. Despite this, we were still able to collect a robust dataset representative of the LTACH setting.

Fourth, because of the difference in total scale scores of the SLUMS and GOT-Cog, sensitivity and specificity testing could not be conducted, and a bootstrap distribution method had to be used to develop cutoff criteria for the GOT-Cog. This heavily relies on the population data, so our datasets from 98 participants dictated what the bootstrap distribution looked like and, therefore, the end cutoff scores. Using these

distributions, we applied clinical expertise to the final decision with regard to which side of the distribution ranges to set the recommended cutoff scores. For example, if the analysis indicated a cutoff of 20, we then had to choose whether to set the criteria below 20 (0–19), or whether 20 should be included in that range (0–20) of the proposed cutoff.

The fifth limitation of the study was that rater reliability testing was not formally completed for either the SLUMS or the GOT-Cog before the validity testing was conducted. Participating occupational therapists did view the recommended Veterans Affairs–produced training video in preparation of the study, as well attended an in-service on how to administer the SLUMS and GOT-Cog; no formal rater reliability was completed. We chose to assess validity testing first, because the accuracy of the measure was initially valued over precision.

Finally, participants were not given a formal neuropsychological assessment to establish whether they had any underlying cognitive disorders. Rather, these exclusion criteria were determined through chart review of the participants’ past medical history. Furthermore, without such neuropsychological assessment, the GOT-Cog can only be used as a screening tool to indicate the presence of cognitive domains of concern and whether further testing is required.


Implications for Occupational Therapy Practice

This study indicates that the GOT-Cog is a valuable cognitive screening tool that can be used to screen cognition in the inpatient LTACH setting, identify potential cognitive domains of concern, and indicate whether a patient may require further evaluation by neuropsychology or speech therapy for further evaluation of a specific cognitive domain. This study has the following implications for occupational therapy practice:

- The results of this screening tool are intended to guide the occupational therapy plan of care; assist with discharge planning; and flag the need for additional services, such as speech therapy and neuropsychology, or further targeted assessments for specific cognitive domains of concern.
- This measure will provide therapists with a functionally based screen to capture any areas of concern and assist in treatment planning.
- This measure should be used in conjunction with other performance-based measures.
- GOT-Cog scores positively correlated with SLUMS scores, all shared domains were significantly correlated between the GOT-Cog and SLUMS, and all unique domains showed significant correlations to the total score of each assessment.

Conclusion

To our knowledge, the GOT-Cog is the first standardized occupational therapy cognitive screen specifically

designed for the inpatient LTACH setting. The purpose of GOT-Cog is to assist in the comprehensive evaluation of cognitive domains relevant to occupational therapy treatment planning in conjunction with other performance-based measures, as well as assessment of cognition in relation to functional tasks, through performance with ADLs and IADLs. This randomized crossover controlled study indicates that the GOT-Cog displays significant correlations to the SLUMS, demonstrating strong criterion and construct validity. Going forward, we plan to recruit inpatients admitted to our LTACH setting to evaluate the interrater reliability, intrarater reliability, and responsiveness of this new screen. 

Acknowledgments

We sincerely thank the members of the Gaylord Hospital Occupational Therapy Department who assisted with data collection, including Alexandra Maneen, Anne Walczak, Bradley Fletcher, Emily Zuckerman, Heidi Fagan, Jaclyn Lavigne, Joseph Giannelli, Katherine Zimmerli, Lauren Rescsanski, Madeline Murgatroyd, Marcia Brassard, Megan Palmer, and Stacey Melillo. Thank you all for your help on this project! We also thank and recognize Rachel Jano for her efforts in assisting with data curation. Last, a subset of these data was presented at the 2023 Annual Conference of the American Congress of Rehabilitation Medicine in Atlanta, GA.

References

- American Occupational Therapy Association. (2013). Cognition, cognitive rehabilitation, and occupational performance. *American Journal of Occupational Therapy*, 67(6, Suppl.), S9–S31. <https://doi.org/10.5014/ajot.2013.67S9>
- American Occupational Therapy Association. (2014, October 6). *The IMPACT act: A step toward post-acute care reform*. <https://www.aota.org/practice/practice-essentials/payment-policy/medicare1/medicare-improving-medicare-post-acute-care>
- American Occupational Therapy Association. (2021, November 24). *Role of occupational therapy in assessing functional cognition*. <https://www.aota.org/practice/practice-essentials/payment-policy/medicare1/medicare-role-of-ot-in-assessing-functional-cognition>
- Faul, M., Wald, M. M., Xu, L., & Coronado, V. G. (2010). *Traumatic brain injury in the United States: Emergency department visits, hospitalizations, and deaths, 2002–2006* (p. 101). Centers for Disease Control and Prevention. <https://stacks.cdc.gov/view/cdc/5571>
- Giles, G. M., Edwards, D. F., Baum, C., Furniss, J., Skidmore, E., Wolf, T., & Leland, N. E. (2020). Making functional cognition a professional priority. *American Journal of Occupational Therapy*, 74, 7401090010. <https://doi.org/10.5014/ajot.2020.741002>
- Hrdlicka, H. C., Meise, E., Corbett, J., Meyer, A., & Grevelding, P. (2024). Content validity testing of a novel cognitive screen, the Gaylord Occupational Therapy Cognitive (GOT-Cog), to improve inpatient occupational therapy treatment planning. *American Journal of Occupational Therapy*, 78, 7801345020. <https://doi.org/10.5014/ajot.2024.050306>
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Noyes, E. T., Major, S., Wilson, A. M., Campbell, E. B., Ratcliffe, L. N., & Spencer, R. J. (2023). Reliability and factor structure of the Saint Louis University Mental Status (SLUMS) examination. *Clinical Gerontologist*, 46, 525–531. <https://doi.org/10.1080/07317115.2022.2120446>
- Okie, S. (2005). Traumatic brain injury in the war zone. *New England Journal of Medicine*, 352, 2043–2047. <https://doi.org/10.1056/NEJMp058102>
- Portney, L. G., & Gross, K. D. (2020). Chapter 10: Concepts of measurement validity. In L. G. Portney (Ed.), *Foundations of clinical research: Applications to evidence-based practice* (4th ed., p. 127–140). F. A. Davis.
- Saint Louis University. (n.d.). *SLU Mental Status Exam*. <https://www.slu.edu/medicine/internal-medicine/geriatric-medicine/aging-successfully/assessment-tools/mental-status-exam.php>
- Tariq, S. H., Tumosa, N., Chibnall, J. T., Perry, M. H., & Morley, J. E. (2006). Comparison of the Saint Louis University Mental Status Examination and the Mini-Mental State Examination for detecting dementia and mild neurocognitive disorder—A pilot study. *American Journal of Geriatric Psychiatry*, 14, 900–910. <https://doi.org/10.1097/01.JGP.0000221510.33817.86>
- Wolf, T. J., Barco, P. P., & Giles, G. M. (2019, November 18). *Functional cognition: Understanding the importance to occupational therapy*. American Occupational Therapy Association. <https://www.aota.org/publications/ot-practice/ot-practice-issues/2019/functional-cognition>

Emily Meise, MS, OTR/L, is Occupational Therapist, Inpatient Occupational Therapy, and Clinical Research Liaison, Milne Institute for Healthcare Innovation, Gaylord Specialty Healthcare, Wallingford, CT.

John Corbett, BA, is Clinical Research Coordinator, Milne Institute for Healthcare Innovation, Gaylord Specialty Healthcare, Wallingford, CT.

Amanda Meyer, MS, OTR/L, is Occupational Therapist, Hartford Healthcare, Hartford, CT, and Occupational Therapist, Inpatient Occupational Therapy, Gaylord Specialty Healthcare, Wallingford, CT.

Pete Grevelding, MSPT, is Executive Director, Milne Institute for Healthcare Innovation, Gaylord Specialty Healthcare, Wallingford, CT.

Henry C. Hrdlicka, PhD, is Director of Research, Milne Institute for Healthcare Innovation, Gaylord Specialty Healthcare, Wallingford, CT; hhrdlicka@gaylord.org