Objective. We designed a simple rating instrument, the Occupational Therapy Task Observation Scale (OTTOS), to facilitate evaluation and documentation of patient performance during occupational therapy task groups and to improve the communication between occupational therapists and other treatment team members.

Method. After analysis of other rating instruments, a preliminary version of OTTOS was designed, extensively field tested, and further modified. Interrater reliability was determined, and validity was tested against three established rating instruments: the Bay Area Functional Performance Evaluation, the Comprehensive Occupational Therapy Evaluation Scale, and the Milwaukee Evaluation of Daily Living Skills.

Results. The final version of OTTOS contains two parts, 10 items for evaluation of specific task functions and 5 items for rating general behavior. Field use demonstrated that the scale successfully tracked changes in functional capacity and included most facets of patient function evaluated during task groups. Use of OTTOS required minimal training, and scoring required less than 2 min for each patient. The correlation between the scores of experienced occupational therapists was high (.92 for the total scores). The correlation between OTTOS and the other rating instruments ranged from .880 to .340; the highest correlations, as expected, were with test subscales that most closely resembled OTTOS.

Conclusion. Preliminary data indicated that OTTOS offers a reliable and valid method for rapidly rating the function of patients with psychiatric illness attending task groups. In addition, use of OTTOS improved the communication between occupational therapists and other health care providers, facilitated the education of occupational therapy students, and met the documentation requirements of third-party payers.

Task groups play a fundamental role in the contribution of occupational therapy to contemporary psychiatric evaluation and treatment (Bradlee, 1993; Mosey, 1986). Tasks performed in the group can be adjusted to the functional level of individual patients, which may vary premorbidly from patient to patient and change rapidly during the course of psychiatric illness. Further, the groups permit a single occupational therapist to evaluate the capacities of a number of patients rapidly and simultaneously; such efficiency is critical given the financial restraints that have led to reductions in the length of inpatient psychiatric treatment and decreases in the intensity of routine occupational therapy provided to inpatients. Most importantly, by effectively communicating observations made during task groups to other treatment team members, the occu-
The American Journal of Occupational Therapy

A 75-year-old woman with severe depression was admitted to an inpatient psychiatric unit. A 2-month outpatient course of a standard antidepressant medicine had not led to a reduction in symptoms. Her physician decided to add a second antidepressant medicine, hoping for an additive effect. After several days, nursing staff members commented that the patient seemed somewhat drowsy but was complaining less of depression and anxiety than on admission. The occupational therapist, however, noted a marked deterioration in the patient's performance during task group: poor ability in following directions, decline in concentration, less socialization with other patients, poor coordination, and decline in the quality of her work. The occupational therapist reported this deterioration to the physician, who suspected that the patient was complaining less to the nursing staff members because she was becoming delirious. After confirming this diagnosis, the physician stopped all psychoactive medicines.

Over the next week, nursing staff members noted an increase in the patient's complaints of depression and anxiety, but the occupational therapist observed an improvement of task skills back to the level observed at admission. The physician used this information to conclude that the delirium had resolved and prescribed a different type of antidepressant. After 2 weeks, nursing staff members noted continued complaints of depression and anxiety, but the occupational therapist observed a gradual improvement in task performance as the patient engaged more readily in tasks with improved decision-making ability, increased independence, and better problem-solving skills. On the basis of this information, the physician decided to continue the new medicine, and the treatment team members pointed out to the patient that she was doing better, even though she was not yet feeling better. After another 2 weeks, task performance continued to improve, and the patient described a consistent improvement in her mood.

The occupational therapist's role in the evaluation of patients with psychiatric disorders is to evaluate function. The occupational therapy evaluation may provide a vastly different picture of a patient's clinical state than the specific symptoms reported by the patient. For example, as noted by observant clinicians for at least a century, recovery from depression is often apparent in a patient's appearance, socialization, and performance of activities of daily living before he or she reports feeling less depressed (Kraepelin, 1921/1989). These improvements, which may not be apparent in a standard psychiatric interview with a patient, can readily be detected by an experienced occupational therapist observing the patient's social and functional skills during a task group. For patients with delirium and dementia, subtle changes in cognition, attention, or level of consciousness may only be apparent while observing them perform a specific functional task over time. Disorders of personality may be partially manifested by a marked difference between actual task performance and the level of dysfunction claimed by the patient. The challenge for the occupational therapist is communicating these observations to other treatment team members, tracking a large number of patients simultaneously, quantifying changes, and documenting evaluations for reimbursement.

Several instruments exist that partially address these requirements. The Bay Area Functional Performance Evaluation (BaFPE) (Bloomer & Williams, 1979; Williams & Bloomer, 1987), a sophisticated and well-validated scale for evaluation of performance of a variety of index tasks (Houston, Williams, Bloomer, & Mann, 1989; Klyczek & Mann, 1990; Thibeault & Blackmer, 1987), is limited in that it requires an hour or more of attention to one patient to administer and score. The American Association on Mental Deficiency Adaptive Behavior Scale is oriented toward evaluation of patients with cognitive impairment and is not designed to assess task group performance or day-to-day functional change (Miller, 1972). The Comprehensive Occupational Therapy Evaluation Scale (COTES) (Brayman, Kirby, Misenheimer, & Short, 1976), a lengthy list of items that can be applied to a task group setting, is cumbersome to use in practice, has not been extensively validated, and offers a fairly limited range for scoring individual items. Two older rating scales (Clark, Koch, & Nichols, 1965; Wolff, 1961) do not use the language of contemporary occupational therapy.

We created the Occupational Therapy Task Observation Scale (OTTOS) to address the need for a modern, simple, quantitative, and rapid method of evaluating task performance that would enhance communication between the occupational therapist and other members of a psychiatric treatment team. In this article, we describe OTTOS and the procedures for its use, present preliminary data pertaining to reliability and validity, and discuss its impact on the practice of psychiatric occupational therapy.

Development and Use of OTTOS

A list of descriptors of task performance was developed, with terminology largely derived from the Model of Human Occupation, that at face value, covered a wide range of performance components observed during task groups. To form OTTOS, the descriptors were organized into a 16-item task behavior section and a 5-item general behavior section. Existing instruments were used as models for both form and content. After 12 occupational therapists (each with at least 1 year of experience) piloted OTTOS almost daily for approximately 1 year, the scale was revised and incorporated into a hospital-approved form (see Figure 1). In addition to OTTOS proper, the form includes a comment section.

Each section of OTTOS is scored from 0 to 100, and each item within a section is scored from 0 (maximal
### Task Behavior Subscore = \[
\frac{\sum \text{Scores}}{100}
\]

### General Behavior Subscore = \[
\frac{\sum \text{Scores}}{100}
\]

---

**Figure 1.** The Occupational Therapy Task Observation Scale (OTTOS)© as formatted for chart documentation. Note: Printed with permission from Johns Hopkins Medical Institutions, Department of Psychiatry and Behavioral Sciences.
dysfunction) to 10 (no evidence of dysfunction). Proper use of OTTOS requires the maintenance of an absolute standard for each item instead of a standard relative to an individual patient’s best and worst level of function. The disadvantage to the use of an absolute standard is that the scores of patients who have either extremely high or extremely low levels of function will remain relatively fixed, which will be discussed later in the article. However, because all patients are scored on the same scale, the final score is easy for nonoccupational therapists to interpret, and comparison of OTTOS scores to scores from other rating instruments is possible.

Reliability

Though OTTOS was principally devised for longitudinal evaluation by a single therapist, preliminary data pertaining to interrater reliability were obtained. Two experienced occupational therapists who did not participate in the development of OTTOS simultaneously observed psychiatric inpatients attending occupational therapy task groups (N = 25; with 19 rated on both sections) and then independently rated each patient with OTTOS. Correlation coefficients were .89 for task behavior scores, .76 for general behavior scores, and .92 for the total scores (p < .0005 for each comparison). Reliability was also evaluated by comparing an experienced occupational therapist’s OTTOS ratings with those of an occupational therapy student after both had observed 10 psychiatric day hospital patients attending a task group. These scores were not highly correlated (.33 for task behavior scores, -.11 for general behavior scores, and .37 for total scores; p > .05 in each case). The small sample size makes any conclusion tentative but suggests that rater experience influences OTTOS scores.

Validity

Face validity was evaluated through the routine use of OTTOS by occupational therapists working with the various patient populations admitted to the psychiatric service of Johns Hopkins Hospital in Baltimore, Maryland. Criterion validity was explored by comparing OTTOS scores with BaFPE and COTES scores obtained for the same patients, on the same day, by the same occupational therapist (OTTOS vs. COTES: N = 24; OTTOS vs. BaFPE: N = 21). All subjects were hospitalized inpatients who gave informed consent for participation.

To validate OTTOS against a reliable and validated measure of functional living skills, it was compared with the Milwaukee Evaluation of Daily Living Skills (MEDLS), a tool used to establish the level of assistance required by persons with chronic mental illness in performing standard living skills (Leonardelli, 1988). For each of 62 patients hospitalized in a short-stay community psychiatry unit, OTTOS scores (mean scores of one task group near admission and a second task group at discharge) and the scores from nine MEDLS subtests (taken mid hospitalization) were obtained and correlated. Pearson product–moment correlations were used to test the significance of correlation for all comparisons (see Table 1).

OTTOS scores correlated highly with scores for all COTES scales. The highest correlation was between OTTOS task behavior and COTES task scores (.880), and the lowest was between OTTOS general behavior and COTES task scores (.592). The correlation of the total scores of the two scales was .826.

Correlation between OTTOS and BaFPE scores was somewhat lower. The highest correlation was between the scores of the BaFPE cognitive subscale and the OTTOS task behavior section (.726), whereas the lowest correlation was between the scores of the OTTOS general behavior section and the BaFPE performance subscale (.340). The correlation between OTTOS and MEDLS scores was similar to those between OTTOS and BaFPE, ranging from .459 to .624.

The high correlation between OTTOS scores and COTES scores was expected because the COTES was also intended to rate function during task groups. The lower correlation between OTTOS scores and BaFPE and MEDLS scores suggests that task group behavior reflects somewhat different functional processes than the life skills measured by MEDLS or the specific task skills measured by BaFPE. The OTTOS task behavior section was more internally reliable and more closely correlated with all three instruments than the general behavior section. This difference between the two sections most likely arises from the more numerous and more specifically functional capacities addressed in the task behavior section. The relatively high correlation between the BaFPE cognitive subscale and the OTTOS task behavior section indicates, as might be expected, that the specific task behavior skills evaluated with OTTOS are rooted in fundamental cognitive skills directly evaluated by more sophisticated instruments such as the BaFPE.

Discussion

OTTOS has now been used extensively by psychiatric occupational therapists at Johns Hopkins Hospital. The scale has proven useful in several ways. First, OTTOS has improved the ability of the occupational therapists to report the clinical progress of individual patients to other treatment team members. Therapists using OTTOS found that the structure of the scale directed their attention to
particular aspects of a patient’s functional status, such as the ability to follow directions or plan ahead, and facilitated the evaluation of change in functional capacity over time. The wide range in scoring for each item (0–10) enabled users to note subtle changes in patient functional capacity, an important feature in documenting the often slow progress of persons with severe mental illness.

Second, OTTOS has proven valuable for record keeping. The scale takes only 1 or 2 min to complete and reduces the need for long narrative descriptions, though the hospital form into which OTTOS is incorporated allows for a narrative supplement to the scale items. Documentation time has been reduced by an estimated 50%. In addition, third-party payers favor the type of detailed, quantitative summary of patient problems that the OTTOS provides.

Third, OTTOS has been a helpful teaching device. The availability of a simple but reasonably comprehensive list of task behaviors has increased the speed with which occupational therapy students learn to focus and describe their observations of patients in task groups. Reference to OTTOS categories and scores during discussions of patient diagnosis and progress enables clinicians without formal training in occupational therapy (such as physicians, nurses, and social workers) to develop a greater awareness of its role in patient evaluation.

Certain limitations of OTTOS are important to recognize. The reliability and validity data presented in this article are preliminary. Different forms of analysis, such as test–retest reliability, as well as other measures of validity will need to be considered. Also, clinical experience indicated that OTTOS was most useful in evaluating the progress of patients functioning at lower levels. For patients functioning at higher levels, particularly those with affective disorders, scores were so tightly clustered at the top of the range that progress was difficult to follow. This ceiling effect partially stems from the limitations of craft-based task groups; our experience is that evaluation of patients with high function requires more demanding and individualized tasks. Our preliminary experience is that OTTOS is useful for gauging patient progress in activities requiring high functional capacity, such as cooking, typing, and individualized projects related to academic studies or employment. However, in patients whose functional capacity is not detectably impaired by psychiatric illness, the purpose of occupational therapy must shift away from a combination of evaluation and treatment toward treatment alone. Finally, the scale cannot replace the narrative descriptions of experienced therapists; we view OTTOS as supplementary and as facilitating such descriptions. Limitations notwithstanding, OTTOS appears to be a clinically valuable tool for quantitatively rating the functional capacity of patients participating in task groups.

Acknowledgments

We thank Monique Howard, OTR, and Andrea Schutz, OTR, for assistance with data collection.

The authors request that those readers who wish to use OTTOS send (or E-Mail: rmargol@welchlink.welch.edu) written descriptions of the purpose for which they will use the scale as well as any comments or suggestions for improving the scale.

References


---

Table 1

<table>
<thead>
<tr>
<th>OTTOS Subscale</th>
<th>COTES Subscale</th>
<th>BaFPE Subscale</th>
<th>MEDLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td>Interpretation</td>
<td>Task</td>
</tr>
<tr>
<td>Task behavior</td>
<td>.763*</td>
<td>.798*</td>
<td>.890*</td>
</tr>
<tr>
<td>General behavior</td>
<td>.607**</td>
<td>.710*</td>
<td>.592*</td>
</tr>
<tr>
<td>Total</td>
<td>.740*</td>
<td>.814*</td>
<td>.792*</td>
</tr>
</tbody>
</table>

Note. High function on the COTES is scored as a low number; hence, the absolute value of COTES correlations were used for clarity. OTTOS = Occupational Therapy Task Observation Scale, COTES = Comprehensive Occupational Therapy Evaluation Scale, BaFPE = Bay Area Functional Performance Evaluation, MEDLS = Milwaukee Evaluation of Daily Living Skills.

* p < .0005. ** p < .005. *** p < .025. **** p < .05. ***** p > .05.


