The Test of Orientation for Rehabilitation Patients: Interrater Reliability

Jean C. Deitz, Vicky S. Tovar, Deborah W. Thorn, Clara Beeman

Key Words: brain injuries • cognition • reliability of tests • research design • tests, by title, Test of Orientation for Rehabilitation Patients

The Test of Orientation for Rehabilitation Patients (TORP) (Deitz, Beeman, & Thorn, in press) was developed for use with patients with brain injuries in inpatient rehabilitation settings. It was designed to assess orientation to person and personal situation, place, time, schedule, and temporal continuity. Interrater reliability for the TORP was examined with the use of 34 brain-injured and 35 non-brain-injured patients. Two occupational therapists trained in administering the TORP, as specified in the test manual (Dietz et al., in press), served as the examiners. One therapist administered and scored the test while the second therapist observed and scored the test for the same subject. Intraclass correlation coefficients, used as indexes of reliability for the scoring of the total test and subtests, ranged from .89 to 1.00 for the non-brain-injured group and from .94 to .99 for the brain-injured group. These findings suggest that an occupational therapist can reliably score the TORP for patients both with and without brain injuries.

Jean C. Deitz, PhD, OTR/L, FAOTA, is Associate Professor, Department of Rehabilitation Medicine, RJ-30, University of Washington, Seattle, Washington 98195.

Vicky S. Tovar, MS, OTR/L, at the time of this study, was a graduate student in the Department of Rehabilitation Medicine, University of Washington, Seattle, Washington.

Deborah W. Thorn, MS, OTR/L, is a Mental Health Professional, Community Home Health Care, Seattle, Washington.

Clara Beeman, MOT, OTR/L, is an Occupational Therapy Supervisor, Harborview Medical Center, Seattle, Washington.

This article was accepted for publication January 24, 1990.

Persons with brain injuries often display symptoms of confusion and disorientation in the acute stages following their injuries. The degree to which these symptoms persist varies greatly, but patients with brain injuries usually continue to display orientation deficits at the time they enter the rehabilitation phase of their treatment. The rehabilitation of such patients can be seriously hampered by their difficulty in recalling and using information regarding person, place, time, schedule, and temporal continuity (Bond, 1975; Haberman, 1982; Levin, Grossman, Rose, & Teasdale, 1979). Disorientation can impair function in the areas of self-care, work, and leisure activities. Occupational therapists often help patients with brain injuries compensate for the effects of disorientation on their functional abilities by evaluating the patient's orientation, devising treatment strategies, and assessing the effectiveness of these strategies for improving function.

One obstacle to the assessment of orientation and the evaluation of treatment effectiveness has been the absence of a tool to provide a thorough, objective, reliable, and valid measure of orientation for persons being treated in inpatient rehabilitation settings. The Test of Orientation for Rehabilitation Patients (TORP) (Deitz, Beeman, & Thorn, in press) was designed to meet the need for such an assessment. It is a standardized, fixed-schedule questionnaire that uses a face-to-face, open-ended format with an auditory recognition task to help separate deficits in information retrieval from those in orientation. The TORP has a standardized method of test administration and an objective scoring system. It consists of 46 items divided into five domains of orientation: (a) Person and Personal Situation, (b) Place, (c) Time, (d) Schedule, and (e) Temporal Continuity. It yields a total test score, which reflects general orientation, and separate domain scores. The therapist may find this test to be valuable in making diagnoses, planning treatment, evaluating treatment effectiveness, and measuring orientation for research purposes.

The purpose of this study was to evaluate the interrater reliability and interrater agreement of the TORP. Although the TORP was designed for clinical use with patients with brain injuries, we studied patients both with and without brain injuries, because evidence of instrument reliability must be demonstrated for all populations for which the test will be used.

The following research questions were addressed separately for a brain-injured and a non-brain-injured group of rehabilitation patients:

1. What is the test form preparation agreement for the TORP?
2. What are the interrater reliability and the in-
ttering agreement of the total test scores on the TORP.

3. What are the interrater reliability and the in­
terrater agreement of the domain scores on the TORP?

The last question was addressed separately for each of the five domains of orientation.

Method

Sample

Occupational therapists in five inpatient rehabilitation units in Seattle identified patients for both the brain-injured and non-brain-injured groups according to predetermined criteria. Criteria for inclusion in ei­ther group were that the subject be at least 14 years of age, have no history of hospitalization for a psychiatric disorder, have no history of an insidious neurological disease such as Alzheimer, and have been on a reha­bilitation unit for a minimum of 7 consecutive days. For the brain-injured group, the therapists were asked to refer only those patients whom they considered to be appropriate candidates for an orientation test. These were patients for whom the therapists’ assess­ment goals were either to define the extent of disori­entation or to rule out the possibility of orientation deficits. Additional criteria for inclusion in the brain­injured group were a diagnosis of brain injury and no significant aphasia, as determined by a speech and language pathologist. An additional criterion for in­clusion in the non-brain-injured group was a medical condition or history not associated with brain injury.

The sample for this study consisted of 34 patients with brain injuries (18 men, 16 women) and 35 pa­tients without brain injuries (14 men, 21 women). The subjects with brain injury ranged in age from 18 to 86 years (m = 54 years); the subjects with no brain injury ranged in age from 15 to 87 years (m = 49 years). The median number of days between admission to a rehabilitation unit and TORP administration was 14 for patients in the brain-injured group and 11 for patients in the non-brain-injured group. The diag­nostic categories in the brain-injured group were closed-head injury, cerebrovascular accident, cerebral aneurysm or hemorrhage, brain abscess, brain cancer, and gunshot wound. The diagnostic categories in the non-brain-injured group were orthopedic problems, neuromuscular problems, arthritis, coronary disease, vascular disease, amputation, and Addison disease.

The subjects were included in the study in the order in which they became available for testing and agreed to participate. We tried to include subjects from a variety of age groups and diagnostic categories. Subjects in certain age and diagnostic groups were no longer accepted after several subjects in those groups had already been tested (e.g., female subjects over the age of 70 years with cerebrovascular accident).

Instrument

The TORP was developed in the Department of Reha­bilitation Medicine at the University of Washington in Seattle. It grew from a need for a thorough, standard­ized, reliable, valid, and objective test of orientation for rehabilitation patients with brain injuries. The content validity of the TORP was examined at both the item and domain levels in a previous study (Thorn & Deitz, 1990), the results of which provided evidence of content validity.

The TORP is divided into five domains. The first domain, Person and Personal Situation, relates to “knowing who one is, knowing who significant persons in one’s environment are, and knowing inform­ation pertaining to oneself and one’s background. This includes order of events in one’s past” (Deitz et al., in press). Questions in this domain include “How old are you?” and “What kind of work were you doing before your (illness/injury)?” (Deitz et al., in press).

The second domain, Place, is defined as “know­ing significant information about one’s current envi­ronment and being aware of distance and direction. Current environment includes locations relating to the hospital or rehabilitation center and to one’s present home address” (Deitz et al., in press). Examples of Place questions are “What (city/town) are you in now?” and “What direction is (Canada/United States) from here?” (Deitz et al., in press).

The third domain, Time, is defined as being “able to identify the current time including: (a) time of day, (b) day of the week or month, (c) month, (d) season, and (e) year” (Deitz et al., in press). Examples of questions in this domain are “What season is it now?” and “About what time do you think it is now?” (Deitz et al., in press).

The fourth domain, Schedule, relates to “being aware of the time and sequence of appointments and discrete events in one’s current routine or schedule” (Deitz et al., in press). Questions from this domain include “What time is your next physical therapy appoint­ment?” and “Which meal will be your next meal?” (Deitz et al., in press).

The last domain, Temporal Continuity, focuses on “being aware of the passage of time or time along a continuum” (Deitz et al., in press). Questions from this domain include “What season comes after spring?” and “How many weeks, months, or years has it been since you last (attended school/worked at a job/worked as a homemaker)?” (Deitz et al., in press).

The TORP’s format is unique in that each test item is written both as an open-ended question and as
Table 1
Reliability of Test Form Preparation for the Non-Brain-Injured (n = 35) and the Brain-Injured (n = 34) Groups

<table>
<thead>
<tr>
<th>Domain</th>
<th>Agreement (%)</th>
<th>Partial Agreement (%)</th>
<th>Disagreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person and Personal Situation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>83.8</td>
<td>13.2</td>
<td>3.0</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>83.6</td>
<td>12.5</td>
<td>3.9</td>
</tr>
<tr>
<td>Place</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>99.0</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>99.5</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>96.7</td>
<td>3.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>96.1</td>
<td>3.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Schedule</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>94.8</td>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>92.2</td>
<td>3.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Temporal Continuity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>84.1</td>
<td>14.3</td>
<td>1.6</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>81.9</td>
<td>10.9</td>
<td>7.1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-brain-injured</td>
<td>90.6</td>
<td>8.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Brain-injured</td>
<td>89.9</td>
<td>7.0</td>
<td>3.1</td>
</tr>
</tbody>
</table>

an auditory recognition task. The latter is used only if the patient fails to respond correctly to the open-ended question. The recognition tasks were included in the test to help distinguish information retrieval deficits from storage deficits. Before administering the TORP, the test administrator enters the correct answers in the boxes provided for the items. The auditory recognition task relates to the same content as the open-ended question. It is followed immediately by a randomized list of three distracters and the correct response. An alternate distracter is provided when the correct response is one of the distracters.

Preparation for TORP administration requires approximately 15 min for the therapist to fill in correct responses from the medical chart and usually another 10 min to call a patient's family member for information not in the chart. The time required to administer the TORP varies from less than 5 min for patients who are not disoriented to 30 min for patients who are very disoriented.

The TORP was designed so that therapists, rehabilitation professionals, or other professionals, working in pairs, can learn administration and scoring procedures by completing a training module contained in the TORP manual (Deitz et al., in press). The training process involves a series of steps. First, the therapists read and study the TORP test manual. Second, working in pairs, they complete training examples for preparing the test form. These training examples consist of simulated medical records, test forms, and correct answer keys. Third, the two therapists, using examples included in the training module, practice test administration through role-playing. Fourth, the two therapists independently prepare the test forms for a minimum of six patients and then compare their results. Fifth, they administer the TORP to six patients, alternating who both administers and scores the TORP and who only scores the TORP. Twice during the training process, the examiners are required to complete a written competency test. For all activities, the training module describes minimum levels of competence required before the therapist can proceed with the next phase of training.

Procedure

Examiners and examiner training. Two occupational therapists with experience working with rehabilitation patients with and without brain injuries were hired as examiners for this study. As specified in the TORP manual, both had some knowledge of the general principles and procedures of standardized testing (Deitz et al., in press), but neither had experience with the TORP. Their training, prior to the commencement of data collection, consisted of completing the training module for the TORP. As specified in the TORP manual, each examiner completed the training while paired with another occupational therapist working in the local community. These community therapists met the same selection criteria as those serving as examiners for the study and likewise had no experience with the TORP.

The two examiners trained separately and, once data collection commenced, were instructed not to discuss with each other the test form preparation or the test administration and scoring processes. This

Table 2
Domain-Corrected Scores and Total-Test-Corrected Scores for Examiners A and B for the Non-Brain-Injured (n = 35) and the Brain-Injured (n = 34) Groups

<table>
<thead>
<tr>
<th>Domain</th>
<th>Non-Brain-Injured</th>
<th>Brain-Injured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person and Personal Situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.91</td>
<td>1.68</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.90</td>
<td>1.69</td>
</tr>
<tr>
<td>Place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.95</td>
<td>1.72</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.96</td>
<td>1.73</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.90</td>
<td>1.50</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.90</td>
<td>1.50</td>
</tr>
<tr>
<td>Schedule</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.73</td>
<td>1.15</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.72</td>
<td>1.15</td>
</tr>
<tr>
<td>Temporal Continuity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.90</td>
<td>1.50</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.91</td>
<td>1.50</td>
</tr>
<tr>
<td>Total test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner A</td>
<td>1.89</td>
<td>1.56</td>
</tr>
<tr>
<td>Examiner B</td>
<td>1.89</td>
<td>1.57</td>
</tr>
</tbody>
</table>
was done to determine the interrater reliability of two therapists who had trained without interacting with each other.

Data collection. The testing process commenced with each examiner independently preparing a test form for the subject before testing. Correct completion of the test form is critical to the testing process because answers to most of the TORP items are unique to each patient. Information was obtained from the subjects' medical record as well as from staff and family members.

Using the test form that each had prepared, one therapist administered and scored the test while the second therapist observed and scored the test for the same subject. The subjects in the brain-injured and non-brain-injured groups were assigned alternately to each examiner in the order in which the referrals to each group were received. Each of the examiners administered the test to approximately half of the subjects in the brain-injured group and half of the subjects in the non-brain-injured group.

As specified in the TORP training module (Deitz et al., in press), the examiners and their training partners, using patients who were not part of this study, rechecked their scoring agreement after 30 subjects had been tested. This was done to control for observer drift, which is described by Kazdin (1982) as the tendency of observers to change the way in which they score data.

Data Analysis

We used four steps to examine the consistency with which two examiners could score the TORP. First, agreement in preparing the test form was analyzed with a point-by-point percentage agreement. Agreement, partial agreement, and disagreement were considered. Agreement occurred when both examiners recorded the same possible correct answers for an item. Partial agreement occurred when at least one of the correct answers recorded by the two examiners matched. Disagreement occurred when the correct answers recorded by the two examiners were totally different. For example, for the item, “What is the last name of a medical doctor you see here?” (Deitz et al., in press), if both Examiner A and Examiner B indicated that the subject sees Dr. Smith and Dr. Jones, then an agreement would be recorded. If Examiner A indicated that the subject sees Dr. Smith and Examiner B indicated that the subject sees only Dr. Jones, then a partial agreement would be recorded. If Examiner A indicated that the subject sees Dr. Jones and Examiner B indicated that the subject sees Dr. Smith, then a disagreement would be recorded.

Second, descriptive statistics (i.e., means, medians, standard deviations, and low and high scores) for each examiner for each of the five domains and for the total test were calculated. Corrected scores were used for all analyses, because these scores are adjusted according to the number of items administered to an individual. A person's score, therefore, is not compromised if up to five items cannot be administered because the items are inappropriate for the individual or because the correct answers are not available. Corrected scores can range from 0.00 to 2.00. The former occurs when a person incorrectly answers
all of the items administered; the latter occurs when a person correctly answers all of the open-ended questions for the items administered.

Third, intraclass correlation coefficients were used as indexes of interrater reliability for each of the five domains of the TORP as well as for the total test. The one-way analysis of variance (ANOVA) model was used, as opposed to the two-way random effects model, because although all raters rated all subjects, the selection of raters was not random (Tinsley & Weiss, 1975). Because the raters were not selected randomly, the one-way ANOVA model provides a more conservative estimate of reliability. This model assumes that any differences between raters will contribute to unreliability in the measures (Armstrong, 1981; Shrout & Fleiss, 1979). Mean squares were obtained from the ANOVA output generated by the reliability subprogram provided by SPSS Version 7–9 (Hull & Nie, 1981). In addition, although two raters were used in this study, the calculated intraclass correlations reflect the expected reliabilities of a single judge's ratings.

Last, interrater agreements between examiners on both domain and total test scores were determined. Tinsley and Weiss (1975) advised that this be done in addition to an examination of interrater reliability, because it is possible to have one without the other. Specifically, the magnitudes of difference between examiners on both domain and total test scores were calculated to determine the actual differences in the scores assigned by the two examiners.

## Results

The reliability of test form preparation for the non-brain-injured and brain-injured groups is shown in Table 1. The percentage of agreement for both groups for the domains ranged from 81.9% to 99.5%, with the lowest agreements in the domains of Person and Personal Situation and Temporal Continuity.

Descriptive statistics for domain scores and total test scores for both groups are shown in Table 2. Out of a possible score of 2.00 for each of the domains and for the total test, the means for the subjects in the non-brain-injured group ranged from 1.72 to 1.96. In contrast, the means for the subjects in the brain-injured group ranged from 1.15 to 1.73. For both groups, the means for Examiners A and B never differed by more than 0.01, and there was no evidence that one examiner consistently scored higher than the other examiner.

The intraclass correlation coefficients, used as indexes of reliability, were high for all domains and
for the total test, ranging from .89 to 1.00 for the non-brain-injured group and from .94 to .99 for the brain-injured group (see Table 3). Because the data were not normally distributed, probability values were not reported.

Table 4 shows the magnitudes of difference between the domain-corrected scores of Examiners A and B for the non-brain-injured and brain-injured groups. The magnitudes of difference between the total test scores are shown in Table 5.

### Discussion

The data from the present study suggest that an occupational therapist who has completed the training module for administering and scoring the TORP can reliably score the TORP for patients with and without brain injuries. Intraclass correlation coefficients were uniformly high for both groups for all five domains and for the total test.

Close examination of the magnitudes of difference between scores assigned by the two examiners provides further insight into the consistency with which two examiners can score. More than 90% of the cases had a difference of less than 0.05 in the scores assigned for the total test by the two examiners. In contrast, for the domains, although there was perfect agreement for most of the subjects, the scores sometimes differed by as much as 0.33. This occurred for both groups in the domain of Schedule and for the brain-injured group in the domain of Time. Because each domain has only six items, disagreement on even one item resulted in the scores differing by 0.33. For the domains that have more items and for the total test, however, the scoring of only one item differently does not affect so greatly the respective domain and total scores assigned by the two examiners. The clinician, therefore, can have more confidence in the scores on the total test and on the larger domains than in the scores on the smaller domains.

A unique feature of this study was the examination of the reliability of test form preparation. Interrater reliability did not appear to be substantially affected by less-than-perfect agreement in the filling out of the test forms. We believe this to be because in most cases, when agreement was not perfect, it was partial. Partial agreement tells us that when more than one answer is acceptable, Examiner A and Examiner B overlapped in identifying acceptable responses, but that one examiner had one or more answers designated as correct than did the other examiner. This does not appear to affect interrater reliability, probably because both examiners identified the most obvious correct answers. For example, in the case of a patient who sees a physiatrist and a neurologist regularly and a gynecologist occasionally, Examiner A might answer the question, "What is the last name of a medical doctor you see here?" (Deitz et al., in press), by listing all three physicians, whereas Examiner B might list only the first two. Because the patient is required to give the name of one physician only, the patient's response would likely be to give the name of one of the first two physicians. In such a case, the difference in test form preparation would not affect interrater reliability.

In the domain of Temporal Continuity, the slightly lower agreement for test form preparation appeared to reflect minor math errors, thus suggesting the importance of one's rechecking all calculations.

A weakness of the present study was the use of only two raters, which was done because we believed it to be inappropriate and potentially disruptive to bring more than two examiners into the testing room with a patient. To counteract this weakness, we considered using videotapes and multiple raters. Such an approach, however, would not have provided information on the feasibility of administering and scoring.
the test simultaneously. The results, therefore, would
have been more generalizable to other therapists but
less generalizable to the typical test administration
situation in the clinical setting.

Conclusion

The results of this reliability study suggest that ther­
apists can score with consistency patients’ perfor­
mance on the TORP. Future research efforts are
needed to further examine interrater reliability and
agreement and to examine test-retest reliability and
construct and criterion-related validity.

Acknowledgments

We wish to thank Kari Einset and Suzanne Richmond, who
collected data, the clinicians who identified potential sub­
jects and referred them to this study, the patients who par­
ticipated in this study, and Mike Trevisan, who provided
statistical consultation.

This study was supported in part by Research Grant No.
G00830076 from the National Institute of Handicapped Re­
search, Department of Education, Washington, DC.

References

Armstrong, G. (1981). The intraclass correlation as a
measure of inter-rater reliability of subjective judgments. 
Nursing Research, 30, 314-315, 320A.

Bond, M. R. (1975). Assessment of the psychosocial
outcome after severe head injury. Ciba Foundation Sym­

Deitz, J., Beeman, C., & Thorn, D. (in press). Test of
Orientation for Rehabilitation Patients (TORP). Seattle:
Health Sciences Center for Educational Resources, Univer­
sity of Washington.

Haberman, B. (1982). Cognitive dysfunction and social
rehabilitation in the severely head-injured patient. Journal
of Neurosurgical Nursing 14, 220-224.

procedures and facilities for releases 7-9. New York:
McGraw-Hill.

University Press.

Levin, H. S., Grossman, R. G., Rose, J. E., & Teasdale,
G. (1979). Long-term neuropsychological outcome of

tions: Uses in assessing rater reliability. Psychological Bul­
netin, 86, 420-428.

of the Test of Orientation for Rehabilitation Patients. Occu­

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliabil­
ity and agreement of subjective judgements. Journal of
Counseling Psychology, 22, 358-376.

"Technologically" Speaking

Stay on top of the technological issues affecting occupa­
tional therapy by adding these resources to your professional
library!

Just Published!

Technology Review '90

This collection of papers offers the major technological
presentations made at AOTA's 70th Annual Conference in
New Orleans. Focusing on the value of technology as it re­
lates to occupational therapy, Technology Review '90 in­
cludes summaries and illustrations of both low- and
high-tech applications.

$12.00 AOTA member, $16.00 non-member

Technology Review '89

Twenty leading authors share fresh perspectives on current
and future technological innovations in this collection of
papers presented at AOTA's 69th Annual Conference in
Baltimore.

$17.00 AOTA member, $23.00 non-member

Order both Technology Review '90 and Technology Review
'89 together and save!

$22.00 AOTA member, $27.50 non-member

Applications of Technology for the Disabled—
Introduction and Overview

This videotape provides a comprehensive overview of
technology, demonstrating how technology can assist
individuals with mobility, communication, cognitive tasks,
environmental control, and computer access.

$35.00 AOTA member, $43.00 non-member

Order now with MasterCard or VISA! Call 1-800-THE­
AOTA (AOTA members), 1-800-654-5584 (Md. mem­
ers), or (301) 948-9626 (non-members).