Developing a Context-Based Performance Measure for Persons With Schizophrenia: The Test of Grocery Shopping Skills

Edna Hamera, Catana E. Brown

Key Words: environment • skills assessment

Objective. This article describes how authenticity and directness, desired characteristics of performance measures, were applied to the development of a context-based Test of Grocery Shopping Skills (TOGSS) for persons with schizophrenia.

Method. The steps used in developing the measure included interviewing consumers with schizophrenia to identify issues in grocery shopping, conceptualizing how authenticity and directness could be applied to shopping performance, and selecting grocery items to be used in the test. The two forms of the TOGSS were administered to 26 persons with schizophrenia or schizoaffective disorders to evaluate reliability (stability, equivalence, interrater) and validity (convergent, generalizability) of the TOGSS.

Results. The correlations between the two forms of the TOGSS over two different testing periods were significant, ranging from .64 to .83. Subscale scores were moderately correlated (R = .52. to .94) with a similarly constructed test of drugstore shopping.

Conclusion. A systematic method can be used to develop a context measure of performance. The TOGSS has beginning evidence of reliability and validity. With further study, the test will be useful in assessing the independent living skill of grocery shopping in persons with psychiatric disorders.


The current emphasis on cost-effective outcomes requires demonstration of the effect of interventions on occupational performance. Performance measures are the most direct method for evaluating the effectiveness of occupational therapy. For numerous reasons, such as cost and efficiency, many measures assess performance in simulated situations, effectively neglecting an important component of functioning: the context of performance. Assessing performance in context increases the possibility that results are a valid reflection of performance in daily life.

Assessment in context is particularly relevant for persons with schizophrenia, who often experience a variety of information processing deficits. These deficits can include (a) difficulty in processing subtle and complex environmental cues (Cornblatt & Keilp, 1994); (b) difficulties in screening out distractions (McDowd, Filion, Harris, & Braff, 1993); (c) deficiency in short-term memory, affecting the ability to attribute meaning to stimuli (Knight, 1993); and (d) diminished ability to use spatial and temporal consistencies in the environment (Hemsley, 1996). The presence of these deficits make it difficult for persons...
with schizophrenia to generalize skills learned in simulated situations to the real world.

One of the more challenging skills for this population is grocery shopping because this activity requires selecting a desired product at an affordable price in an environment designed to promote impulse buying. Given the essential nature of this cognitively complex skill and the “noisy” context where it occurs, grocery shopping was selected as the skill and persons with schizophrenia as the target population for assessing performance in context.

This article describes the process of developing a reliable and valid performance measure of grocery shopping, including steps used in development of the test and the evaluation of its reliability and validity. The conceptual framework guiding the development of the measure and the application of validity to performance measures provide the background of the test.

Context-Based Frameworks in Occupational Therapy

The role of the environment, a term used interchangeably with context in the occupational therapy literature, has always been central in assessment and skill training in occupational therapy. However, understanding of the concept became more sophisticated with the emergence of new conceptual frameworks. An example of how environment is used in a conceptual model is illustrated in the Occupational Adaptation Model (Schkade & Schultz, 1992). This model describes occupational adaptation as a process that occurs in transition periods when persons must adjust to changes in the environment or in themselves. Another framework, the Ecology of Human Performance (Dunn, Brown, & McGuigan, 1994), suggests that a person cannot be understood without a complete understanding of the person’s context. This model includes an explicit description of intervention approaches that can be directed at the person, the environment, or the task.

The same movement is taking place in occupational therapy outside of this country. For example, the Person–Environment–Occupation Model (Law et al., 1996), originating from Canada, proposes that the dynamic transactions of person, environment, and occupation come together as occupational performance. The importance of improving the fit of person, occupation, and environment is accentuated. Closely related is the Person–Environment–Occupational Performance Model (Christiansen & Baum, 1997), which recognizes the complexity of the person–occupation–environment transaction. This model also attends to the role of motivation and meaning in occupational performance.

These four conceptual frameworks call on occupational therapists to integrate context in daily practice. This task would include the development of assessments that are contextually relevant; that is, assessments in which occupational performance is evaluated through approaches that incorporate actual performance of skills in natural settings.

Previous studies indicate that context influences assessment performance. Two studies of older adults that used the Assessment of Motor Process Skills found that participants performed better in a familiar, home environment as opposed to a clinical setting (Nygård, Bernspång, Fisher, & Winblad, 1994; Park, Fisher, & Velozo, 1994). Performance on two tasks of the Kohlman Evaluation of Living Skills resulted in different outcomes in clinical versus natural environments (Brown, Moore, Hemman, & Yunek, 1996). Other investigators studying persons with dementia found that indirect evaluation using the subject as reporter resulted in overestimated performance, whereas indirect evaluation using an informant (e.g., family member) as reporter resulted in underestimated performance (Karagiozis, Gray, Sacco, Shapiro, & Kawas, 1998). Thus, a relevant environment becomes an important component of valid evaluation.

Validity of Performance Measures

With the increased popularity of performance measures in education and research, psychometricians are beginning to define how validity applies to performance measures. According to Messick (1994), performance measures need to be established using the concepts of authenticity and directness. Authenticity refers to behaviors that are relevant to real-world experience. This involves identifying and adequately representing the construct assessed by the performance measure, including both the process and the quality of the domain measured. Evidence for authentic validity includes

1. substantive validity (i.e., whether the construct[s] are measured relevant to the population);
2. content validity (i.e., whether the sampling of skills represents the concept[s]); and
3. structural validity (i.e., the fidelity of the scoring structure to the structure of the construct).

Messick (1994) referred to directness as the degree to which performance reflects real life, which means a minimization of method variance that is not related to the construct(s) being measured. For example, if fine motor coordination is part of an assessment of grocery shopping, but performance does not depend on fine motor coordination, then variance resulting from fine motor coordination is not related to the construct of shopping. Directness includes external validity; that is, convergent and discriminative evidence from comparing performance on the measure with similar and dissimilar constructs and methods, as well as demonstrating the use of the measure. Generalizability, another dimension of directness, involves examining consistency across populations, settings, and tasks.

Developing the Test of Grocery Shopping Skills

Our aim in developing a test of grocery shopping was to evaluate the effectiveness of an intervention. Because assess-
ment of baseline shopping performance was desired, the development of two different forms for pretesting and posttesting was necessary. Developing the test of grocery shopping involved the following steps:

1. Interviewing consumers with schizophrenia to identify the issues that they have in grocery shopping
2. Conceptualizing how authenticity and directness could be applied to grocery shopping performance
3. Selecting grocery items for the test.

Interviews of Consumers
Ten persons with schizophrenia from a community-based program were interviewed to elicit the issues they had with grocery shopping. The open-ended interview questions concerned (a) where and how they typically obtained groceries, (b) what issues were important to them in deciding what to purchase, and (c) what foods they commonly use. Most of the participants received help with transportation to the store, and a few reported needing help with finding items. Participants reported that price and finding foods that were filling were the most important factors in their decision about what to buy. Two outcomes emerged that were relevant in measuring grocery shopping: getting the items that one wants for the lowest price (i.e., accuracy), and finding the items in an efficient manner (i.e., efficiency).

Application of Authenticity and Directness
Messick's (1994) concept of authenticity was applied in developing the test by assessing performance in a natural setting (a grocery store) and by designing the test so that it mirrored real-life performance. Specifying the constructs essential to shopping supported content validity. Accuracy, which focuses on the quality of performance, and efficiency, which focuses on the procedures for shopping, guided the development of the measure and guarded against a narrow interpretation of the task. At the same time, the task was limited to actually finding the items in the store because preliminary shopping activities (e.g., deciding what to purchase, planning the shopping trip) do not occur in the grocery store. Evaluation of these activities would involve developing a separate test that could be applied where the participant lives. The inclusion of items commonly used by the population and of items that represent the major food groups contributed to authenticity (i.e., substantive validity). Structural validity was maintained by keeping the operationalization of accuracy and efficiency broad enough to encompass the constructs.

To meet the requirement of directness, a decision was made not to limit the strategies the person uses in finding items (e.g., how one navigates the store, how one uses the assistance of others). Method variance was minimized, however, by giving uniform instructions, by evaluating whether participants grasped the salient points that were being assessed before measuring their performance, and by specifying which grocery items to obtain. The amount of time it took to complete the test was measured from when the participant got a shopping cart to when he or she entered the checkout line. A grocery store that was not in the area where participants usually shop was selected to control different degrees of familiarity with the store and to reduce error variance. Participants were told that they could keep the grocery items they found to minimize differences in motivation.

Selecting Grocery Products for the Test
A list of 27 grocery products was compiled to generate two parallel forms of the test, each with 10 grocery items; one form of the test to assess baseline performance and the other to assess change with intervention. Three criteria were used in selecting grocery products for the test:

1. Items commonly used by consumers with schizophrenia
2. Items that offer a number of choices in flavors or added ingredients and sizes
3. Items that were distributed throughout the store

Twenty consumers with schizophrenia from a community-based mental health program were randomly assigned to shop at one of two mid-size grocery stores that were outside the area in which they lived. The 27-item Test of Grocery Shopping Skills (TOGSS) was administered to each participant at the store. While participants looked at the instructions, the research assistants read them aloud:

You will be given a list of 27 items that you will need to locate and select. Pay close attention to the size of the item on the list and to the brand that is the lowest cost for the size. I will accompany you into the store and take notes but will not be able to answer questions. I will be timing you. This does not mean you need to rush but try not to waste time. You may keep the items you find that are on the list.

Participants were given a clipboard with the grocery shopping list attached and a pen. The time when the participants entered the store to when they entered the checkout line was recorded. The research assistants used a scoring sheet to record whether participants selected the correct items, the correct size, and the lowest price. The brand that had the lowest price for each item was listed on the scoring sheet and reviewed weekly by the research assistants. Entries on the scoring sheet were validated using the checkout receipt generated at the checkout line.

The total accuracy score was the sum for the number of correct items, number of items at the correct size, and number of items at the lowest price. Of the three accuracy indexes, finding the lowest price for the item was the most difficult for participants. Out of a possible range of 0 to 27, the mean on this index was 16.9. The mean for finding the correct item (23.6) and the correct item size (23.95) were
higher. Time in minutes from entering the store to entering the checkout line, one index of efficiency, ranged from 40 to 105 min, with a mean of 68.20 min. Measuring distance with pedometers, another index of efficiency, was problematic because the pedometers did not function consistently.

The original idea of using two stores to avoid participants learning from the environment also proved problematic. The total accuracy score differed between the two stores (M = 60.8 for Store A, 69.9 for Store B). Efficiency measured by time also differed between the two stores (M = 72.9 min for Store A, 61.1 min for Store B). These findings suggested that the environment of the two grocery stores was not comparable. Finding two comparable stores did not appear feasible, so a decision was made to use one store for both forms of the test. Anticipating that the test would be administered before and after an intervention of several weeks, we decided that familiarity with the store would have limited effect on the test scores.

Specific grocery items were selected for the two forms of the test using a three-step process:

1. Identification of items with difficulty between 20% and 65%
2. Examination of item difficulty correlations to find equivalent pairs of items
3. Division of the 20 items selected into 10 pairs, one set for Form 1 and the other for Form 2, so that each form had items distributed throughout the store

The final TOGSS has 10 grocery items on each of the two forms. Each item is given 1 point for correct performance on each of three indexes: (a) selecting the correct item, (b) selecting the correct size, and (b) selecting the lowest price for the item. Total accuracy is the sum of the three indexes with a possible range of 0 to 30. Efficiency is measured by the number of minutes it takes to select all 10 items.

Evaluation of the TOGSS

A test–retest design was used to evaluate consistency between the two forms of the TOGSS. Convergent validity and generalizability were evaluated by comparing scores on a similar test of shopping and participants’ responses to an interview about their grocery shopping habits.

Participants and Procedure

Twenty-six different persons with schizophrenia or schizoaffective disorders who were recruited from two community support programs participated in the study. Of the 26, 18 were women and 8 were men; 18 were white, 4 were African-American, 2 were Latino, and 2 were biracial. They ranged from 22 to 57 years of age, with an average age of 40.5 years. The average gross monthly income (including housing and food subsidies) was $655. Twenty participants lived independently, 4 lived alone with some in-home assistance, and 2 lived with relatives.

At the initial contact, participants were interviewed about their grocery shopping habits, and information was obtained about demographic and treatment variables. In a subsequent session, at one of the same grocery stores used to identify items for the TOGSS, either Form 1 or Form 2 of the TOGSS was administered. Participants were given the same instructions used when piloting the TOGSS. However, to make sure that participants understood the instructions, an additional question was added. Participants were asked to repeat the instructions and were reminded of the important points if they could not recall them. At the second testing, 25 to 35 days later, the form of the TOGSS not given initially was administered.

To examine convergent validity and generalizability of grocery shopping to shopping in another setting, the Test of Drug Store Shopping Skills (Wise & Brown, 1997) was also administered at a different time of the same day at the second testing or on another day of the same week. The Test of Drug Store Shopping skills requires participants to find 15 items from a list, and it is scored for accuracy and efficiency in a similar manner as the TOGSS.

To ensure that the research assistants were administering and scoring the TOGSS uniformly, they were taught by the investigators, practiced with each other, and were evaluated as they administered the TOGSS to each other. To evaluate consistency in scoring over time, one investigator was present and independently completed the scoring sheet eight times when Form 1 of the TOGSS was administered and seven times when Form 2 of TOGSS was given to participants in the study.

Analysis

Content analysis was used to categorize participants’ responses to an item that asked “how do you decide what to buy” on the semistructured interviews about grocery shopping habits. Answers were rated on a 3-point scale from no strategy to some use of strategy to use of effective strategy. Intercode reliability assessed between two independent raters was 85.7%.

The reliability and validity analyses measuring the relationship between one measure and another were analyzed using Pearson product-moment correlation. Gender differences were examined using dependent sample t tests.

Results

Reliability

The descriptive statistics for each of the three indexes of Form 1 and Form 2 of the TOGSS are shown in Table 1. The relationship of performance on the two forms of the TOGSS over two different testing periods provides a coefficient of stability (test–retest) and equivalence (two forms). This coefficient is usually lower than a coefficient that measures only one aspect of reliability, either stability or equiv-
Results showed men had higher total accuracy scores than women on Form 1 ($M = 24.6$ vs. $M = 21.5$) and on Form 2 ($M = 23.4$ vs. $M = 20.2$) but were less efficient on Form 1 ($M = 21.5$ vs. $M = 19$) and Form 2 ($M = 20.2$ vs. $M = 19.9$). These differences, however, were not significant.

### Discussion

The development of the TOGSS illustrates the challenge of developing reliable and valid context-based performance measures. Authenticity and directness, characteristics of performance measures (Messick, 1994), guided the development of the test. In striving for authenticity, grocery shopping concerns of the population were elicited from interviews, and the findings were incorporated into the test. For example, the ability to find the lowest priced item was included because this was a major issue of the population. A measure of efficiency was also included because members of this population often have others waiting for them when grocery shopping and because inefficiency can compound the frustration level of this cognitively taxing task. Some degree of authenticity was sacrificed to maintain reliability. Participants did not shop for items of their own choosing, and the items that were listed included a specificity that would not be present on typical grocery lists (e.g., “14-oz can of powdered cleanser”).

In striving for directness, the TOGSS requires participants to carry out the actual task of grocery shopping in a natural environment of a grocery store. This is a very direct test. However, one measure of efficiency—distance—was problematic. The shuffling gait of some of the participants, along with pedometer malfunctions, contributed to data collection problems. A different method for assessing this aspect of efficiency is being evaluated (e.g., have research assistants wear the pedometer).

Empirical evidence of authenticity and directness involved evaluating the test’s reliability and validity. The results show that the TOGSS may be a useful measure of an important independent living skill. Stability was achieved in that participants did not learn from the initial test, suggesting that the TOGSS can be used to evaluate whether participants improved after an intervention.

Beginning evidence of validity is supported by correspondence of the TOGSS with a measure of shopping in another setting. A correlation in the expected direction, but not significant, was found between scores on the TOGSS and self-report of shopping strategies. A study using a larger sample or a more detailed assessment of strategies might achieve significance between grocery shopping strategies and grocery shopping performance.

Although gender differences in ability to grocery shop might be expected in a general population, this was not anticipated in persons with schizophrenia, and no differences were found between men and women. This finding may

### Table 1

<table>
<thead>
<tr>
<th>Test Form</th>
<th>M</th>
<th>SD</th>
<th>Obtained R–Possible R</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form 1–Item</td>
<td>8.15</td>
<td>2.31</td>
<td>1–10</td>
<td>.69</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Form 2–Item</td>
<td>8.56</td>
<td>1.93</td>
<td>0–10</td>
<td>.83</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Form 1–Size</td>
<td>8.92</td>
<td>2.00</td>
<td>1–10</td>
<td>.60</td>
<td>.001</td>
</tr>
<tr>
<td>Form 2–Size</td>
<td>8.22</td>
<td>1.97</td>
<td>1–10</td>
<td>.60</td>
<td>.001</td>
</tr>
<tr>
<td>Form 1–Price</td>
<td>5.35</td>
<td>2.06</td>
<td>1–8</td>
<td>.60</td>
<td>.001</td>
</tr>
<tr>
<td>Form 2–Price</td>
<td>5.70</td>
<td>2.28</td>
<td>1–9</td>
<td>.60</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note: N = 26.

### Table 2

<table>
<thead>
<tr>
<th>Test and Subscales</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form 1 of TOGSS and Test of Drug Store Shopping Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item subscale</td>
<td>.81</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Size subscale</td>
<td>.94</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Price subscale</td>
<td>.44</td>
<td>.025</td>
</tr>
<tr>
<td>Total accuracy</td>
<td>.91</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time</td>
<td>.52</td>
<td>.006</td>
</tr>
<tr>
<td>Form 1 of TOGSS and Test of Drug Store Shopping Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item subscale</td>
<td>.76</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Size subscale</td>
<td>.87</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Price subscale</td>
<td>.67</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Total accuracy</td>
<td>.86</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time</td>
<td>.65</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Note: N = 26.
have resulted from reduced power from the small sample size, but the small differences that did exist were not necessarily in the expected direction. Findings on gender differences in schizophrenia were inconsistent, but any differences tend to be attenuated when using stable community-based clients (Goldstein, 1993; Sood, Baker, & Bleedin, 1996).

Future research to support convergent and discriminant validity is needed for the TOGSS. A study of the relationship between grocery shopping performance and cognitive tests that measure the cognitive skills needed for grocery shopping would contribute to convergent validity. Improvement in performance on the TOGSS after an intervention to improve grocery shopping skills would also support the test’s validity as a measure that can detect change. Although the sample included only persons with schizophrenia, the TOGSS may be useful for other persons with severe and persistent mental illness as well as for persons with other types of brain disorders who have difficulty maintaining concentration and using environmental cues in a setting with many competing stimuli.

The weakness of the TOGSS is its limited focus on the task of grocery shopping without consideration of the planning phase of shopping and completing shopping by purchasing the items. Comprehensive evaluation of grocery shopping skills involves a series of context-based performance measures, starting with meal planning, list preparation, and evaluation of items on hand.

Developing the TOGSS highlighted the inherent challenge of measuring performance in a natural and dynamic setting. During the study, the grocery store that had been chosen changed ownership, which led to some changes in products and product location. Weekly sales occasionally changed what had been the lowest priced item. Weekly evaluation of these changes was the only way to minimize error variance. This dynamic quality of natural settings increases the cost of assessing performance in context. However, evaluation in context may be cost-effective if it is used to screen those persons who need intervention and to allow interventions to be tailored to individual needs.

Conclusions
The importance of performance measures is clear in a profession that focuses on evaluating skills in natural settings. However, it is also important to recognize the inherent challenges of developing reliable and valid performance measures. These challenges include integration of a conceptual framework and application of psychometric concepts relevant to performance measures. These processes can be used in developing other performance-based measures.

The TOGSS described in this article was developed with consideration of the unique characteristics of performance measures; that is, authenticity and directness. Our efforts at establishing equivalence, test–retest reliability, and beginning validity demonstrate that the test is a promising tool for measuring grocery shopping skills.

Acknowledgments
We thank Angela Kolenbrander and Jennifer Peterson for data collection and Chris Wise for development of the Test of Drug Store Shopping. This research was funded by a grant from the American Occupational Therapy Foundation.

References


