Reliability Of A Market Basket Assessment Tool (MBAT) For Use In SNAP-Ed Healthy Retail Initiatives

By: Sarah A. Misyak, PhD1; Valisa E. Hedrick, PhD, RDN2; Ellen Pudney, MS, RDN1; Elena L. Serrano, PhD1; Alisha R. Farris, PhD, RDN3

Abstract
Objective: To evaluate the reliability of the Market Basket Assessment Tool (MBAT) for assessing the availability of fruits and vegetables, low-fat or nonfat dairy and eggs, lean meats, whole-grain products, and seeds, beans, and nuts in Supplemental Nutrition Assistance Program–authorized retail environments. Methods: Different trained raters used the MBAT simultaneously at 14 retail environments to measure interrater reliability. Raters returned to 12 retail environments (85.7%) 1 week later to measure test-retest reliability. Data were analyzed using paired-sample t tests and correlations. Results: No significant differences were found for interrater reliability or test-retest reliability for individual categories (mean differences, 0.0 to 0.3 ± 0.2 points) or total score (mean difference, 0.5 ± 0.4 points and (mean differences, 0.0 to 0.3 ± 0.3 points) or total score (mean difference, 0.8 ± 0.4 points), respectively. Conclusions and Implications: Future steps include validation of the MBAT. A low-burden tool can facilitate evaluation of efforts to promote healthful foods in retail environments.

Reliability of a Market Basket Assessment Tool (MBAT) for Use in SNAP-Ed Healthy Retail Initiatives
Sarah A. Misyak, PhD; Valisa E. Hedrick, PhD, RDN; Ellen Pudney, MS, RDN; Elena L. Serrano, PhD; Alisha R. Farris, PhD, RDN

ABSTRACT
Objective: To evaluate the reliability of the Market Basket Assessment Tool (MBAT) for assessing the availability of fruits and vegetables, low-fat or nonfat dairy and eggs, lean meats, whole-grain products, and seeds, beans, and nuts in Supplemental Nutrition Assistance Program–authorized retail environments.

Methods: Different trained raters used the MBAT simultaneously at 14 retail environments to measure interrater reliability. Raters returned to 12 retail environments (85.7%) 1 week later to measure test-retest reliability. Data were analyzed using paired-sample t tests and correlations.

Results: No significant differences were found for interrater reliability or test-retest reliability for individual categories (mean differences, 0.0 to 0.3 ± 0.2 points) or total score (mean difference, 0.8 ± 0.4 points), respectively.

Conclusions and Implications: Future steps include validation of the MBAT. A low-burden tool can facilitate evaluation of efforts to promote healthful foods in retail environments.

Key Words: assessment, reliability, retail nutrition environment, SNAP-Ed, Dietary Guidelines for Americans

INTRODUCTION
The Supplemental Nutrition Assistance Program–Education (SNAP-Ed), which is funded through the US Department of Agriculture’s (USDA’s) Food and Nutrition Service, is the nutrition education arm of SNAP, the largest government nutrition assistance program in the US. The goal of SNAP-Ed is to “improve the likelihood that persons eligible for SNAP will make healthy food choices within a limited budget.” The SNAP-Ed operates in all 50 states, Washington DC, and US territories, providing evidence-based nutrition and physical activity direct education and comprehensive, multilevel interventions that include policy, systems, and environmental changes where people eat, live, work, learn, shop, and play.

Many SNAP-Ed–implementing agencies carry out healthy food retail initiatives within the shop setting using SNAP-Ed–endorsed programs as guides (eg, Stock Healthy, Shop Healthy, the Minneapolis Healthy Corner Store Initiative, and the Baltimore Healthy Corner Store Initiative). These efforts generally focus on increasing the availability of healthful foods in food retail outlets within local communities. However, storeowners and managers tend to perceive that consumers value convenience over nutritional quality of food items and a higher demand for unhealthy food and beverage items over healthy ones. Researchers and practitioners may experience difficulty convincing store owners to decrease the availability of unhealthy items as part of retail-focused intervention.

Agencies that implement SNAP-Ed are expected to evaluate and report impacts from their work; therefore, low-cost, low-burden, practical environmental audit measures are needed to assess baseline availability and/or changes based on their efforts. Tools exist that are designed to collect detailed information on the availability and pricing of healthy and unhealthy food items, but information on unhealthy items may be of limited use to researchers and practitioners; it takes significant time and resources to complete such information given store owners and managers’ unwillingness to decrease the number of unhealthy food and beverage items that are available.

In an effort to address these current limitations, a simple, lower-burden, standardized evaluation tool was...
created that could be widely used within local and statewide SNAP-Ed retail initiatives. Standardizing the tools used can better inform practice and policy,13 particularly as SNAP-Ed continues to move toward standardized evaluation through the SNAP-Ed Evaluation Framework.14 This newly developed Market Basket Assessment Tool (MBAT), which is available as online Supplementary Data, provides information on the availability and quality of food items that closely align with the 2015–2020 Dietary Guidelines for Americans (DGA) within a variety of retail environments.15 Following Townsend’s16 process for developing reliable, valid, sensitive, and practical measures for SNAP-Ed evaluation, the specific purpose of this study was to investigate the reliability of the MBAT.

METHODS
Instrument Development
A team of experts in nutrition, evaluation, SNAP-Ed, and community outreach conducted a review of literature for tools that assess the retail nutrition environment. A score was assigned based on the number of categories of available foods, the number of varieties of foods within those categories, and the quality of available food. Possible total scores ranged from 0 to 40, with higher scores indicating greater availability, quality, and varieties of items from the different food categories. The scoring structure was developed based on the scoring scheme of an existing audit tool.17 Categories included fresh, frozen, and canned fruit and vegetables; lean meat and fish; low-fat and nonfat dairy and eggs; whole-grain products; and dried beans, nuts, and seeds. The MBAT was pilot-tested for 1 year (2015–2016) for face validity in a variety of retail settings and store types including grocery stores, dollar stores, pharmacies, and supercenters. Researchers with expertise in nutrition who were trained in using other environment audit tools assessed the MBAT for the following criteria: level of difficulty or ease of completion, adequate representation of healthy items, and structure and division of food categories on the tool. Based on feedback received, the MBAT was restructured and color-coded to facilitate collecting accurate information; additional food items found to be missing were added within categories; and food categories were subdivided based on fresh, frozen, and canned options.

Procedures
The researchers recruited SNAP-Ed field staff (extension agents) and undergraduate nutrition students serving as SNAP-Ed interns to use the MBAT via a video webinar lasting 1–2 hours. After training, raters tested the MBAT. The trained raters then recruited and trained community stakeholders from health-focused community coalitions to serve as raters to test the ease of completion by community practitioners. An additional 6 raters were recruited and trained. Twelve raters participated in interrater reliability testing and 7 participated in test-retest reliability testing. After consultation with Virginia Tech Institutional Review Board staff, it was determined that institutional review board approval was not required for this study because human subjects were not involved, as per US Department of Health and Human Services guidelines,18 which state that the Policy for Protection of Human Research Subjects applies only to research involving human subjects.

For interrater reliability testing, 2 trained raters visited a retail outlet at the same time and completed the audit separately (MBAT 1a and 1b). For test-retest reliability, a rater completed an in-store assessment (MBAT 1a); then the same rater completed a follow-up assessment in the same store 1 week later (MBAT 2). Raters were instructed to complete the original assessment (MBAT 1a) and the retest (MBAT 2) on the same day of the week to minimize potential discrepancies owing to store restocking schedules. Stores were unaware of the research study before the store visits, and raters explained their purpose only when questioned by store personnel.

Data Analysis
Statistical analyses were performed using SPSS statistical software (version 24, IBM, Armonk, NY, 2016). The researchers assessed the MBAT for availability and quality scores for the 6 individual food product categories and for total availability, quality, and overall scores. To evaluate interrater reliability, MBAT scores from 2 trained raters who visited the same outlet on the same day were compared using paired-sample t tests and Pearson bivariate correlations (ie, MBATs 1a and 1b). In addition, an intraclass correlation coefficient (ICC) was assessed to determine interrater reliability for total MBAT scores. Similarly, to assess test-retest reliability, MBAT scores measured at the same outlet 1 week apart were compared via paired-sample t tests and Pearson bivariate correlations (ie, MBATs 1a and 2). Significance was set a priori at P ≤ .05.

RESULTS
Raters collected data with the MBAT between May and August, 2016 at a convenience sample of 14 SNAP-authorized retail outlets across Virginia, including rural and urban areas and a variety of store types. Data were collected in 4 pharmacies, 5 grocery stores, and 5 dollar stores. Total MBAT scores for all retail outlets ranged from 11 to 40, demonstrating a diverse range of healthy food availability at the audited retail outlets.

Interrater Reliability
No significant differences were found for responses across MBATs 1a and 1b for individual food product categories (maximum mean difference, 0.3 ± 0.2 points) or total MBAT score (mean difference, 0.5 ± 0.4 points) between raters on the same day. Moderate to strong correlations were demonstrated across all categories and totals (r = 0.76–0.99) (Table 1). The ICC for the total MBAT score demonstrated excellent agreement for reliability between raters (ICC = 0.991; 95% confidence interval, 0.974–0.997; P ≤ .001).

Test-Retest Reliability
Similar to the interrater reliability findings, no significant differences were found for responses in MBATs 1a and 2 for individual food product categories (maximum mean difference,
**DISCUSSION**

The MBAT was developed as a retail environment audit tool for use by SNAP-Ed—implementing agencies to evaluate healthy food availability and affordability in SNAP-authorized retail outlets for multiple store types used by SNAP shoppers. Results showed the MBAT to be reliable in terms of interrater and test-retest reliability, because the strength of all correlations were above the acceptable limits of 0.5–0.7.16,19 This tool can help to fill a gap in currently available tools to assess the retail food environment.

A recent review of environmental measures of food retail outlets by Glanz et al13 identified the Nutrition Environment Measures Survey for Stores (NEMS-S) and a food store survey developed by the USDA Food Assistance and Nutrition Research Program, using the USDA’s Thrifty Food Plan (TFP) as a framework, as the most widely used tools for auditing stores.20 The NEMS-S has historically been considered the reference standard for auditing food retail outlet environments because of its availability to researchers and determined reliability. Although the NEMS-S is a useful tool for comparing healthy and unhealthy food products within that environment, it may be burdensome for some researchers and community-level practitioners in that it requires an extensive amount of time for training (around 9 hours for the entire NEMS-S training), data collection, data cleaning, and analysis. In addition, different adaptations of the NEMS-S must be used in separate retail environment types,17 which makes direct comparisons and standardization difficult across different types of retail settings.

The USDA food store survey can be used to assess availability and affordability in retail outlets based on the TFP.21 This tool provides a standardized method for collecting information on available food variety, whether foods are affordable for SNAP participants and other low-income households, and whether the foods making up a TFP market basket can be purchased at a given retail location.21 Unfortunately, this tool includes information on food items that are not promoted for consumption by low-income individuals and families through SNAP-Ed programming (eg, margarine, sugar, pancake syrup, chocolate drink mix) and was not tested for reliability. Information from the MBAT can be compared with the external

---

**Table 1. Interrater Reliability of Market Basket Assessment Tool (MBAT)**

<table>
<thead>
<tr>
<th>Food Product Categories (Possible Score)</th>
<th>MBAT 1a (mean ± SD)</th>
<th>MBAT 1b (mean ± SD)</th>
<th>Difference (Mean ± SE)a</th>
<th>P</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grains: total score (0–5)</td>
<td>3.9 ± 1.1</td>
<td>3.8 ± 1.3</td>
<td>0.1 ± 0.2</td>
<td>.75</td>
<td>.76</td>
</tr>
<tr>
<td>Fruit: total score (0–10)</td>
<td>4.6 ± 3.9</td>
<td>4.7 ± 3.8</td>
<td>0.1 ± 0.1</td>
<td>.58</td>
<td>.99</td>
</tr>
<tr>
<td>Availability score (0–7)</td>
<td>3.6 ± 2.5</td>
<td>3.6 ± 2.4</td>
<td>0.1 ± 0.1</td>
<td>.58</td>
<td>.98</td>
</tr>
<tr>
<td>Quality score (0–3)</td>
<td>1.1 ± 1.5</td>
<td>1.1 ± 1.5</td>
<td>0.0 ± 0.0</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Vegetables: total score (0–10)</td>
<td>5.6 ± 3.5</td>
<td>5.4 ± 3.3</td>
<td>0.2 ± 0.2</td>
<td>.39</td>
<td>.97</td>
</tr>
<tr>
<td>Availability score (0–7)</td>
<td>4.6 ± 2.1</td>
<td>4.6 ± 2.1</td>
<td>0.0 ± 0.1</td>
<td>1.00</td>
<td>.98</td>
</tr>
<tr>
<td>Quality score (0–3)</td>
<td>1.1 ± 1.5</td>
<td>0.9 ± 1.4</td>
<td>0.2 ± 0.2</td>
<td>.34</td>
<td>.85</td>
</tr>
<tr>
<td>Meat: total score (0–4)</td>
<td>2.8 ± 1.2</td>
<td>2.7 ± 1.2</td>
<td>0.1 ± 0.1</td>
<td>.34</td>
<td>.98</td>
</tr>
<tr>
<td>Availability score (0–4)</td>
<td>2.8 ± 1.2</td>
<td>2.7 ± 1.2</td>
<td>0.1 ± 0.1</td>
<td>.34</td>
<td>.98</td>
</tr>
<tr>
<td>Quality score (−1 to 0)</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Dairy and eggs: total (0–5)</td>
<td>3.7 ± 1.4</td>
<td>3.4 ± 1.4</td>
<td>0.3 ± 0.2</td>
<td>.10</td>
<td>.91</td>
</tr>
<tr>
<td>Availability score (0–5)</td>
<td>3.7 ± 1.4</td>
<td>3.4 ± 1.4</td>
<td>0.3 ± 0.2</td>
<td>.10</td>
<td>.91</td>
</tr>
<tr>
<td>Quality score (−1 to 0)</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Dried beans, seeds, nuts, and nut butters: total score (0–6)</td>
<td>5.6 ± 0.8</td>
<td>5.6 ± 0.6</td>
<td>0.1 ± 0.1</td>
<td>.58</td>
<td>.78</td>
</tr>
<tr>
<td>Total MBAT score (0–40)</td>
<td>26.2 ± 10.8</td>
<td>25.7 ± 10.4</td>
<td>0.5 ± 0.4</td>
<td>.19</td>
<td>.99</td>
</tr>
<tr>
<td>Total availability score (0–34)</td>
<td>24.1 ± 8.0</td>
<td>23.8 ± 7.9</td>
<td>0.3 ± 0.3</td>
<td>.37</td>
<td>.99</td>
</tr>
<tr>
<td>Total quality score (0–6)</td>
<td>2.1 ± 3.0</td>
<td>1.9 ± 2.8</td>
<td>0.2 ± 0.2</td>
<td>.34</td>
<td>.96</td>
</tr>
</tbody>
</table>

*aMean differences according to paired-sample t test; slight differences may be noted from the preceding columns owing to rounding; bThe correlation cannot be computed because the SE of the difference is 0.

Note: Interrater reliability was assessed by comparing responses from 2 trained raters who visited the same outlet on the same day. MBAT 1a reflects responses from first interrater audits and 1b reflects responses from second interrater audits. Pearson correlations are presented.

0.3 ± 0.3 points) or total MBAT score (mean difference, 0.8 ± 0.4 points) over 1 week. Moderate to strong correlations were demonstrated across all categories and totals (r, 0.74–0.99) (Table 2).
Food Product Categories (Possible Score) | MBAT 1a (mean ± SD) | MBAT 2 (mean ± SD) | Difference (mean ± SE) | P | r
--- | --- | --- | --- | --- | ---
Grains: total score (0–5) | 3.8 ± 1.1 | 4.1 ± 0.9 | 0.3 ± 0.2 | .28 | .74
Fruit: total score (0–10) | 4.3 ± 3.9 | 4.6 ± 3.6 | 0.3 ± 0.3 | .34 | .98
Availability score (0–7) | 3.3 ± 2.6 | 3.6 ± 2.2 | 0.3 ± 0.3 | .34 | .95
Quality score (0–3) | 1.0 ± 1.5 | 1.0 ± 1.5 | 0.0 ± 0.0 | b | b
Vegetables: total score (0–10) | 5.3 ± 3.6 | 5.5 ± 3.5 | 0.2 ± 0.1 | .17 | .99
Availability score (0–7) | 4.3 ± 2.2 | 4.5 ± 2.2 | 0.2 ± 0.1 | .17 | .98
Quality score (0–3) | 1.0 ± 1.5 | 1.0 ± 1.5 | 0.0 ± 0.0 | b | b
Meat: total score (0–4) | 2.8 ± 1.2 | 2.7 ± 1.3 | 0.1 ± 0.1 | .34 | .98
Availability score (0–4) | 2.8 ± 1.2 | 2.7 ± 1.3 | 0.1 ± 0.1 | .34 | .98
Quality score (–1 to 0) | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | b | b
Dairy and eggs: total (0–5) | 3.8 ± 1.3 | 4.0 ± 1.3 | 0.2 ± 0.2 | .34 | .90
Availability score (0–5) | 3.8 ± 1.3 | 4.0 ± 1.3 | 0.2 ± 0.2 | .34 | .90
Quality score (–1 to 0) | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | b | b
Dried beans, seeds, nuts, and nut butters: total score (0–6) | 5.5 ± 0.8 | 5.6 ± 0.7 | 0.1 ± 0.1 | .59 | .77
Total MBAT score (0–40) | 25.6 ± 10.9 | 26.4 ± 10.2 | 0.8 ± 0.4 | .06 | .99
Total availability score (0–34) | 23.6 ± 8.2 | 24.4 ± 7.5 | 0.8 ± 0.4 | .06 | .99
Total quality score (0–6) | 2.0 ± 3.0 | 2.0 ± 3.0 | 0.0 ± 0.0 | b | b

aMean differences according to paired-sample t test; slight differences may be noted from the preceding columns owing to rounding. bThe correlation cannot be computed because the SE of the difference is 0.

Test-retest reliability was assessed by comparing responses from 2 different time points at the same outlet (MBATs 1a and 2) (MBAT 1a reflects responses from the first audit and 1b reflects responses from the second audit). Pearson correlations are presented.

Despite the tool’s strengths, SNAP-Ed implementing agencies should consider its scope before using it, because the MBAT was developed to assess healthy food availability in retail environments. Therefore, it does not include beverages besides low-fat or nonfat milk, and thus data on sugar-sweetened beverages cannot be collected. In addition, the tool does not assess snack foods (eg, granola bars) owing to the difficulty of setting guidelines for what constitutes a healthy snack. Unlike the NEMS-S, the MBAT does not allow for price comparison of healthy and unhealthy foods in store and the price does not factor into scores. However, data can be compared across multiple outlet types within communities and used to assess the availability and price of foods against the external standard of the TFP.

Another limitation is that test-retest reliability would be affected by longer intervals between tests owing to a variety of variables such as seasonality, availability of items, and employee turnover. Another limitation is the relatively small sample size of stores from a single state, which may have affected generalizability of the findings.

**IMPLICATIONS FOR RESEARCH AND PRACTICE**

The MBAT has the potential to decrease burden on SNAP-Ed-implementing agencies and community practitioners conducting needs assessments or evaluating programs to improve food retail environments. Completion time will be assessed with validity testing; however, the associated assessment time is hypothesized to be much lower because the MBAT is 4 pages long, compared with the NEMS-S, which is 15 pages long. The MBAT may be useful for assessing regional and geographic differences in healthy food availability. This tool was developed for SNAP-Ed programs to encourage the use of a standardized evaluation of initiatives to promote healthy food purchases in retail, which would allow for aggregation of results to show the impact of SNAP-Ed on retail environments nationwide.

Ongoing and future research includes testing of the MBAT in a larger number of retail settings and in other locations across the country to ensure...
generalizability, testing the reliability of the MBAT in corner and convenience stores, finalizing validity testing, and assessing for instrument sensitivity. A guide to completing a market basket analysis using data collected through the MBAT is also under development for use by SNAP-Ed programs. Adoption of the MBAT by SNAP-Ed–implementing agencies and different community organizations will also be assessed.

The MBAT can document changes in real and perceived availability of healthy food items after interventions and programs. Future research and programming focusing on strategies to sustain increased perceived and real availability of healthy items that align with the DGA is warranted. These initiatives must be acceptable to store owners and managers, and be sustainable from a business perspective.

ACKNOWLEDGMENTS

This work was partially funded by SNAP-Ed, the USDA, Food and Nutrition Service, and by overhead funding from the Department of Human Nutrition, Foods, and Exercise at Virginia Tech.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jneb.2018.01.012.

REFERENCES


