Policy, Practice and Prevention Research Center

P3RC RESEARCH BRIEF
NO. 125
MARCH 2022

Development and Reliability Testing of the Store CheckOUt Tool (SCOUT) for Use in Healthy Checkout Evaluations

LISA M. POWELL[1], YU LI[1], SARAH E. SOLAR[2], ANDREA A. PIPITO[1], ETHAN C. WOLF[2], JENNIFER FALBE[2]

Key Findings

- **The photo-based Store CheckOUt Tool (SCOUT) provides reliable measures of availability, prices, and promotions of food and beverage products sold at store checkouts.**
- **The average percent agreement across all variables was 0.97.**
- **The average kappa statistic for dichotomous and categorical variables was 0.94 (indicating "almost perfect" agreement).**
- **The average intraclass correlation coefficient for continuous variables was >0.99.**

Introduction

The majority of adults and children in the United States (U.S.) consume added sugars and sodium in excess of the *Dietary Guidelines for Americans'* recommended limits: 10% of total energy from added sugars and 2,300 mg/day for sodium.[1] Consumption of added sugars, including from leading sources like sugar-sweetened beverages (SSBs), sweets, and candy,[2,3] is associated with obesity, type 2 diabetes, cardiovascular disease (CVD), and poor dental health.[4,5] Likewise, excess consumption of sodium, and its major sources such as refined grains, chips, and processed meats,[1,6] increases risk for CVD and also stroke.[7,8] Moreover, the majority of adults and children do not consume sufficient amounts of fruits, vegetables, and whole grains.[1]

Given that two-thirds of calories[9] and added sugars[10] and the majority of sodium[6] in the U.S. diet come from food stores, the retail food environment is an important setting for improving diet quality. A key location within stores is checkout lanes, the only area where all shoppers must pass through and a location where food and beverage manufacturers seek contracts to place their products, making it a profitable space for retailers.[11,12] Evidence shows that checkouts are characterized by a high prevalence of SSBs, candy, sweets, and salty snacks,[13] more than one in three shoppers have reported making purchases of checkout items during their last grocery shopping trip,[14] and checkouts are known for impulse purchases.[15,16] Therefore, checkouts are a high priority for nutrition improvement in the retail setting.

In 2020, Berkeley, CA, became the first U.S. jurisdiction to enact a healthy checkout (HCO) policy.[17] The policy, implemented in March 2021, with enforcement effective January 2022, applies to food stores larger than 2,500 sq ft in Berkeley and prohibits the placement of SSBs, artificially sweetened beverages, and foods that contain >5 g added sugars or >200 mg sodium per serving (e.g., candy, sweets, chips, refined grains, and jerky) at checkout.[17] Allowable products at checkout include non-food products, unsweetened beverages, sugar-free gum or mints, fruits, vegetables, nuts, seeds, legumes, dairy, and whole grains.[17]

The photo-based **S**tore **C**heck**OU**t **T**ool (SCOUT) was developed as part of a larger project to evaluate the impact of the Berkeley HCO policy on availability, pricing, and price promotions of food and beverage products found at food store checkouts and areas adjacent to the checkout. To determine the reliability of the SCOUT for assessing product availability, pricing, and promotions, we conducted an inter-rater reliability (IRR) study in the summer of 2021 based on 316 photos of product facings from store checkouts in 60 stores evenly distributed between Berkeley, Oakland, Davis and Sacramento, California. The purpose of this brief is to describe the SCOUT measures and report on the IRR results.

AUTHOR AFFILIATIONS

1. Health Policy and Administration, School of Public Health, University of Illinois Chicago, Chicago, IL
2. Department of Human Ecology, University of California, Davis, Davis, CA

Methods

Instrument and Measures

The photo-based SCOUT, adapted from our prior work,[18-21] was designed to assess availability, pricing, and price promotions of food and beverage products found at food store checkout and in shopping areas adjacent to the checkout, including standalone displays and aisle endcaps. The SCOUT protocol requires that data are captured via photos for each product facing (i.e., place a product is stocked such as a shelf, slot, hook) at three randomly selected checkouts per store (or all checkouts if there are less than three), standalone displays between the checkouts and main aisles, and aisle endcaps near the checkout areas. Data collectors visit each store and take both wide-angled contextual and up-close detailed product photos at each store location so that location characteristics and product details can be coded from photos. Data collectors upload photos to the project's cloud-based drive.

Data on the location and characteristics of the location are coded for each product. The general location of each facing is coded as: checkout, aisle endcap (front, side/edge), product on floor (≤1 ft from aisle endcap), aisle parallel/perpendicular to checkout, and standalone display (>1 ft from checkout and aisle endcap, ≤1ft from checkout, ≤1 ft from aisle endcap). If the product is displayed at checkout, characteristics include presence/absence of a snaking checkout lane and self-checkout, and a categorical variable to classify product location at checkout: left side of the checkout lane (relative to where the consumer stands), right side of the checkout lane, endcap of checkout, endcap of snaking/funnel section of checkout, inside snaking/funnel section of checkout, above the conveyor belt, on the edge of the conveyor belt, below the register (e.g., shelf or display case below register), behind the register, on top of the register, on the counter near the register, basket or bin attached to lane, on the floor, or hanging above the counter. If the product is located on a shelf, regardless of store location, the total number of shelves and the shelf number for the product are coded.

From the product facing photos, the product category is recorded for both non-food/beverage products (e.g., gift card, magazine, office supply, hand sanitizer) and food and beverage products (e.g., fresh fruit, candy, gum, energy drink, water). Under the current protocol, coders select from 168 categories. For food and beverage products, product characteristics are coded including product brand (e.g., Coca Cola), product detail (e.g., Coke Zero), refrigeration (e.g. fridge/cooler, freezer, not refrigerated) and packaging information (e.g., multipack and number of items in multipack). Additionally, for beverages, container type is recorded (e.g., plastic bottle, glass bottle, unclear bottle, can, box/carton, powder, bag/pouch). Data are coded for up to two package size units (e.g., ounces, milliliters, grams) and the corresponding numeric size.

With respect to promotions, products are coded as to whether they are part of a holiday themed display and whether the product itself is a holiday themed product. In terms of price promotions, indicators are recorded for each product to specify the presence of a current promotion, future promotion (e.g., take $5 off your next purchase), and cross-promotion (e.g., a sale that offers a discount on a different product). For products that have a current promotion, sale type is recorded (for up to two sales) for each product. Sale types include: reduced price, quantity discount (reduced price conditional on purchase of a certain quantity), BOGO (buy any number get any number free), store coupon, manufacturer's coupon, instant rebate, and other. The regular price of each item and the sale price for items with a current promotion are also recorded.

Reliability Study Design and Data Analysis

To determine the SCOUT reliability for assessing product availability, pricing, and promotion, an IRR study was conducted based on images taken of 316 product facings at 60 stores. Each image contained one product facing. The facings represented different types of food and beverage (e.g., gum, candy, chips, sugary beverages, water) and non-food/beverage products. IRR was first assessed for product category, which included categories for type of food, beverage, or non-food/beverage item. Next, for products that were coded in the food and beverage categories, IRR was assessed for product brand, product detail, whether the product was refrigerated, container type (for beverages only), multipack information, product size and size unit, whether it was part of a holiday-themed promotion or was itself a holiday-themed product, whether it had a current, future or cross promotion, sale type, product price, and sale price. Two coders who were trained on the protocol coded the images independently.

Kappa statistics were used for the dichotomous and categorical variables. The kappa statistic is a robust measure of IRR for dichotomous or categorical variables and is a chance-adjusted measure of agreement.[22] Kappa statistics in the range of 0.81–1.00 are considered "almost perfect" agreement, 0.61–0.80 are considered "substantial" agreement, 0.41-0.60 are considered "moderate" agreement, 0.21–0.40 are considered "fair" agreement, 0.00-0.20 are considered "slight" agreement, and anything less than 0.00 is considered "poor" agreement.[23] For kappa statistics, we excluded observations for dichotomous and categorical variables

with a highly skewed prevalence of one response category (i.e., > 0.8). One limitation of using kappa statistics is that if the distribution of a variable is highly skewed, the kappa statistic may be low because the chance level of agreement expected is very high.[24] Thus, we only reported kappa statistics for variables that did not have an average prevalence of any response categories across the two coders exceeding 0.8. Two-way random intraclass correlation coefficients (ICCs) were used to assess IRR for the continuous variables. The ICC is only a valid measure of IRR when there is enough variation in the variable being assessed.[25] Percent agreement, the proportion of responses for a given measure where both data collectors agreed, was reported for all variables. Observations with missing values were excluded from this analysis. All data analyses were performed using Stata/SE 15.1.

Conditional sub-questions were assessed for reliability when coders agreed on the relevant parent questions. For analysis of drink package type, observations were included only if both coders agreed the product belonged to a beverage category. For the number of items in a multipack, observations were included only if both coders agreed the product was a multipack. For sale type and sale price variables, observations were included only if both coders agreed the product had a current price promotion. For analysis of package size, the numeric size was assessed only where the package unit information for the product corresponded across coders. Note that for package size, coders were instructed to report up to two size and unit pairs for the size shown on the front of the product packaging. In instances where coders reported different product metrics that were both units of volume or weight (e.g., one coder reported in fluid ounces and the other in milliliters), the reported metrics were converted to a common metric for reliability analyses.

For current sale type measures, coders were instructed to report sale type information for up to two sales for each product. Seven binary variables were derived to reflect each specific sale type from all reported sales, and a reliability analysis was conducted on each of these binary variables. Additionally, because data were highly skewed for many of the sale types, we created a separate composite sale type variable at the level of each sale occurrence based on all reported sales and conducted a reliability analysis on this categorical variable.

For sale price measures, coders were instructed to report both numeric sale price and sale price text. In instances where there was only sale price text available for a product, coders were instructed to report the text, which we subsequently converted to a numeric value (e.g., "2 for $1" would have an equivalent unit sale price of 50 cents). To ensure adequate variation for analyses, an overall ICC was calculated for sale price variables across all sale types.

Results

The IRR results presented in Tables 1 and 2 show excellent agreement. The average percent agreement across all dichotomous, categorical, and continuous measures was 0.97 (not shown in tables), and the kappa statistics and ICC estimates all showed high inter-rater agreement. Among the 316 product facing images assessed from store checkout areas, 258 were for food or beverage products, and 58 were non-food/beverage product categories. The percent agreement for whether a product was classified as a food or beverage versus a non-food/beverage category was 100%. For the 258 food and beverage products, the tables present the IRR results for each of the product characteristics.

Table 1 presents the kappa statistic and percent agreement results for the dichotomous and categorical measures. As noted above, for kappa statistics, outcomes with highly skewed prevalence (i.e., >0.8) due to high expected agreement by chance were excluded; therefore, Table 1 reports fewer kappa compared to percent agreement statistics. For the dichotomous and categorical variables, the average percent agreement was 0.97 with a range from 0.86 to 1. The kappa statistics ranged from 0.86 to 1 with an overall average kappa of 0.94, which indicates "almost perfect" agreement.[23]

The ICC estimates and percent agreement for the continuous price variables are reported in Table 2. The ICC estimates for both the regular and sale price measures were above 0.99, and the percent agreements were ≥ 0.94.

The primary analysis of the SCOUT IRR excluded observations with a missing value. That is, if one data enterer missed a variable (rather than the variable not being applicable to the product) while the other data enterer reported a value for that variable, the two observations were not assessed. However, in a sensitivity analysis, these observations were included and treated as disagreements in computing percent agreement and kappa statistics. Results from the sensitivity analysis found that the average percent agreement across all variables fell from 0.97 to 0.96 (average percent agreement was unchanged for the dichotomous and categorical variables and fell from 0.95 to 0.92 for the continuous variables), and the average kappa statistic remained the same.

TABLE 1 Kappa and Percent Agreement for Key Product Characteristics

Key Characteristics	Sample Size	Percent Agreement	Kappa
MEAN		**0.97**	**0.94**
Product Category: Food or Beverage	316	1.00	n/a
Product Category: Detail	316	0.97	0.97
Product Brand	250	0.94	0.94
Product Detail	249	0.98	0.98
Product Refrigeration	258	0.97	n/a
Drink Product Container Type	37	0.97	0.95
Multipack	257	0.96	n/a
Number of Products in Multipack	32	0.94	0.93
Package Size Unit 1	225	0.98	0.96
Package Size 1	220	0.86	0.86
Package Size Unit 2	159	1.00	1.00
Package Size 2	159	0.87	0.87
Holiday Display	258	0.98	n/a
Holiday Product	258	0.98	n/a
Current Promotion Availability	251	0.95	0.89
Future Promotion Availability	256	1.00	n/a
Cross Promotion Availability	256	1.00	n/a
Sale Type: Reduced Price	79	0.96	0.92
Sale Type: Quantity Discount	79	0.99	0.97
Sale Type: BOGO	79	1.00	n/a
Sale Type: Store Coupon	79	0.95	n/a
Sale Type: Manufacturer's Coupon	79	1.00	n/a
Sale Type: Instant Rebate	79	1.00	n/a
Sale Type: Other	79	0.99	n/a
Sale Type: Composite*	85	0.96	0.94

Notes: n/a indicates not available because the kappa statistic was not computed due to highly skewed prevalence for the outcome. *The Sale Type Composite measure is a summary categorical variable indicating the specific sale type for 85 sales for 79 products. The summary sale type composite variable was not included in the overall measure reported in this table because the overall measure includes the separate binary sale type measures.

TABLE 2 Intraclass Correlation Coefficient (ICC) and Percent Agreement for Key Product Characteristics

Key Characteristics	Sample Size	Percent Agreement	ICC
MEAN		**0.95**	**0.998**
Regular Price	148	0.95	0.999
Sale Price	81	0.94	0.997

Discussion

The photo-based SCOUT provides highly reliable measures for examining retail and product characteristics including food and beverage availability, pricing, and promotions for products sold at and near store checkouts. The results from this IRR study showed that the average percent agreement across all variables was 0.97. The average kappa statistic for the dichotomous and categorical variables was 0.94 (and ranged from 0.86 to 1 indicating "almost perfect" agreement for all measures), and the ICC estimates for both continuous variables were >0.99.

The results from this study suggest that using this photo-based data collection method in store audits is a particularly reliable means for obtaining data on product availability, pricing, and promotions. The overall reliability estimates from this study are consistent with, and in most cases higher than, those that have been reported, on average, for a wide range of measures from store audit tools that rely on in-person coding.[21,26-32] Previous reliability estimates pertaining specifically to characteristics of products available at checkout are limited and, while some found high reliability for checkout indicators (e.g., for GroPromo, the ICCs for number of checkout side or edge displays were 0.87 and 0.92 respectively),[29] in several cases, reliability was relatively low (e.g., for the CX^3, ICC estimates for the presence of healthful and less healthful products next to the checkout counter were 0.47 and 0.67, respectively;[28] for the FEAD-N tool, the kappa statistic for presence of candy or gum at checkout was 0.28).[30]

Further, existing tools that do collect data on products found at checkout usually capture limited information that can only classify products broadly. The SCOUT captures detailed information on every product facing including brand and size which, in turn, allows researchers to undertake nutritional content analysis for each product.

The comprehensive data collection and high reliability of the SCOUT make it an ideal tool for assessing availability, pricing, and promotions of products sold at and near checkout and for evaluating policies aimed at improving the nutritional profile of products found at checkout.

References

1. U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans, 2020-2025*; December 2020. 9th Edition.
2. Bowman S, Clemens J, Friday J, LaComb R, Paudel D, Shimizu M. *Added sugars in adults' diet: What We Eat in America, NHANES 2015-2016.* Food Surveys Research Group (USDA-ARS); October 2019. Dietary Data Brief No. 24.
3. Bowman S, Clemens J, Friday J, Schroeder N, LaComb R. *Added sugars in American children's diet: What We Eat in America, NHANES 2015-2016.* Food Surveys Research Group (USDA-ARS); December 2019. Dietary Data Brief No. 26.
4. Yang Q, Zhang Z, Gregg EW, Flanders W, Merritt R, Hu FB. Added sugar intake and cardiovascular diseases mortality among US adults. *JAMA Intern Med.* 2014;174(4):516-524.
5. Malik V, Hu FB. The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases. *Nat Rev Endocrinol.* 2022;18:205-218.
6. Drewnowski A, Rehm CD. Sodium intakes of US children and adults from foods and beverages by location of origin and by specific food source. *Nutrients.* 2013;5(6):1840-1855.
7. Larsson SC. Dietary approaches for stroke prevention. *Stroke.* 2017;48:2905-2911.
8. Lichtenstein A, Appel L, Vadiveloo M, et al. 2021 Dietary guidance to improve cardiovascular health: A scientific statement from the American Heart Association. *Circ.* 2021;144:e472-e487.
9. Liu J MR, Li Y, Mozaffarian D. Trends in food sources and diet quality among US children and adults, 2003-2018. *JAMA Netw Open.* 2021;4(4):e215262.
10. Drewnowski A, Rehm CD. Consumption of added sugars among US children and adults by food purchase location and food source. *Am J Clin Nutr.* 2014;100(3):901-907.
11. Klein B, Murphy KM. Exclusive dealing intensifies competition for distribution. *Antitrust LJ.* 2008;75(2):433-466.
12. Harris JL, Webb V, Sacco SJ, Pomeranz JL. Marketing to children in supermarkets: An opportunity for public policy to improve children's diets. *Int J Environ Res Public Health.* 2020;17(4):1284.
13. Basch CH, Fera J. Candy, snack food, and soda in the checkout lines of stores selling products for children in New York City. *J Community Health.* 2021;46(5):922-926.
14. Falbe J, White JS, Sigala DM, Grummon AH, Solar SE, Powell LM. The potential for healthy checkout policies to advance nutrition equity. *Nutrients.* 2021;13(11):4181.
15. Cohen DA, Babey SH. Candy at the cash register—a risk factor for obesity and chronic disease. *NEJM.* 2012;316(15):1381-1383.
16. Cohen DA, Babey SH. Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obes Rev.* 2012;13(9):766-779.
17. City of Berkeley. Ordinance NO. 7,734-N.S. Adding Chapter 9.82 to The Berkeley Municipal Code "Healthy Checkout". 2020. https://www.cityofberkeley.info/Clerk/City_Council/2020/10_Oct/Documents/2020-10-13_Item_02_Ordinance_7734.aspx
18. Falbe J, Lee MM, Kaplan S, Rojas NA, Ortega Hinojosa AM, Madsen KA. Higher sugar-sweetened beverage retail prices after excise taxes in Oakland and San Francisco. *Am J Public Health.* 2020;110(7):1017-1023.
19. Solar SE, Denyer L, Falbe J. Food & Beverage Environmental Audit Tool & Surveillance System. 2020; Healthy UC Davis. https://healthy.ucdavis.edu/funded-projects/completed/food-beverage-assessment.
20. Lee JM, Falbe J, Fanshel RZ. Beverage Logo Advertising and Product Mix Availability on the UC Berkeley Campus. 2020; Berkeley Food Institute. https://food.berkeley.edu/from-the-field/beverage-logo-advertising-and-product-mix-availability-on-the-uc-berkeley-campus/.
21. Li Y, Leider J, Pipito AA, Pugach O, Zenk SN, Powell LM. *Development and reliability testing of a food store observation form for use in beverage tax evaluations.* Chicago, IL: Policy, Practice and Prevention Research Center, University of Illinois Chicago;2018.
22. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37-46.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
24. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37(5):360-363.
25. Koo T, Li M. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-163.
26. Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition Environment Measures Survey in Stores (NEMS-S): Development and evaluation. *Am J Prev Med.* 2007;32:282-289.
27. Hosler AS, Dharssi A. Reliability of a survey tool for measuring consumer nutrition environment in urban food stores. *J Public Health Manag Pract.* 2011;17(5):E1-E8.
28. Ghirardelli A, Quinn V, Sugerman S. Reliability of a retail food store survey and development of an accompanying retail scoring system to communicate survey findings and identify vendors for healthful food and marketing initiatives. *J Nutr Educ Behav.* 2011;43(4):S104-S112.
29. Kerr J, Sallis JF, Bromby E, Glanz K. Assessing reliability and validity of the GroPromo Audit Tool for evaluation of grocery store marketing and promotional environments. *J Nutr Educ Behav.* 2012;44(6):597-603.
30. Izumi BT, Zenk SN, Schulz AJ, et al. Inter-rater reliability of the Food Environment Audit for Diverse Neighborhoods (FEAD-N). *J Urban Health.* 2012;89(3):486-499.
31. Rimkus L, Powell LM, Zenk SN, Han E, Ohri-Vachaspati P, Pugach O, Barker DC, Resnick EA, Quinn CM, Myllyluoma J, Chaloupka FJ. Development and reliability testing of a food store observation form. *J Nutr Educ Behav.* 2013;45(6):540-548.
32. Izumi BT, Findholt NE, Pickus HA, Nguyen T, Cuneo MK. Inter-rater reliability of a food store checklist to assess availability of healthier alternatives to the energy-dense snacks and beverages commonly consumed by children. *Child Obes.* 2014;10(3):266-271.

ACKNOWLEDGMENTS

The reliability study presented in this brief was supported by a grant (2020-85774) from Bloomberg Philanthropies' Food Policy Program (www.bloomberg.org). J.F. was also supported by the NIH/NIDDK (K01DK113068). The contents of this publication do not necessarily reflect the views or policies of Bloomberg Philanthropies or the National Institutes of Health.

SUGGESTED CITATION

Powell LM, Li Y, Solar SE, Pipito AA, Wolf EC, Falbe J. Development and Reliability Testing of the Store CheckOUt Tool (SCOUT) for Use in Healthy Checkout Evaluations. Research Brief No. 125. Policy, Practice and Prevention Research Center, University of Illinois Chicago. Chicago, IL. March 2022. https://p3rc.uic.edu/research-evaluation/evaluation-of-food-policies/retail-food-and-beverages/