INTRODUCTION

Effective program evaluation combines two unlikely partners: rigor and feasibility. This pairing is particularly salient in the evaluation of Supplemental Nutrition Assistance Program Education (SNAP-Ed), which operates through the United States Department of Agriculture (USDA). As the education component of the largest federal nutrition assistance program, SNAP-Ed serves SNAP participants and eligibles by providing evidence-based, behaviorally-focused obesity prevention interventions that include direct education; policy, systems and environmental supports; and social marketing.[1] Specifically, interventions targeting youth are a SNAP-Ed priority. Among the six million recipients of SNAP-Ed direct education in 2012, more than two thirds were school-aged youth.[2] Thus schools have become a focal point for delivering SNAP-Ed.[2,3]

Consequently, school-based assessments of children's nutrition and physical activity behaviors are central to understanding SNAP-Ed effectiveness. However, collecting this data can be daunting for states and implementing agencies. On the one hand, schools can provide optimal settings for the evaluation of obesity prevention programs: the audience is captive, class turnover is relatively low, and assessment can be incorporated into curricula.[2,3] Conversely, competing demands on the educational system and lack of top-down support may be powerful barriers to SNAP-Ed programming and evaluation, especially in under-resourced schools.[2,4] A 2013 Institute of Medicine workshop offered educator perspectives on nutrition education that highlighted election politics, inadequate funding, lack of teacher training, and severely limited instruction time as persistent obstacles to obesity prevention in schools.[2] In terms of evaluation, measures perceived as intrusive, costly, or overly burdensome are more likely to be challenged

by school administrators and teachers.[5] Alternatively, questionnaires that are minimally invasive and easy to administer are more readily accommodated.[5-7]

This problem typifies a broader dilemma in SNAP-Ed evaluation: programs should be systematically assessed with what Guthrie, Stommes, and Voichick[5] refer to as a "common core of measures that would give us a basis for comparison," and yet to be successful, the evaluation must be appropriate and feasible at the community or site level.[5-9] In other words, the quintessential SNAP-Ed measure should be easily applied across diverse contexts and still capture high-quality, consistent data for aggregation on state and national levels.

Over the past decade, impressive strides have been made to guide SNAP-Ed evaluation in this direction.[5-9] In 2006, Townsend[8] published a six-stage process for developing accurate, practical measures to assess community-based nutrition education programs targeting low-income audiences. The tiered stages are outlined in Table 1. More generally, the recent national SNAP-Ed Evaluation Framework and associated *Interpretive Guide*[9] provide direction for evaluating SNAP-Ed, including recommended tools to measure nutrition and physical activity behaviors. These measures are valuable in that they can be applied across diverse contexts to assess indicators linked to SNAP-Ed objectives. Moreover, select tools in the *Interpretive Guide* were validated using the Townsend or a similar process. However, well-validated youth measures requiring over 30 minutes and/or the collection of biometric data[10,11] can be difficult to administer in classrooms, while the *Guide*'s shorter, validated measures are limited in scope (e.g., the Beverage and Snack Questionnaire[12]). SNAP-Ed would therefore benefit from one accurate, practical tool that measures both MyPlate and physical activity behaviors for use with comprehensive school-based obesity prevention programs.

Following Townsend's tiered process for tool development, we present the reliability testing phase (Stage 4) for a Kids' Activity and Nutrition Questionnaire (KAN-Q) that is feasible to administer in classrooms and directly aligns with the national SNAP-Ed Evaluation Framework indicators for healthy eating and physical activity.

METHODS

Early Questionnaire Development

Initial questionnaire development spanned stages 1-3 of the Townsend process[8] (Table 1). A team of experts in nutrition, SNAP-Ed, and evaluation selected content domains (Stage 1) by examining nutrition education instruments for school-aged children in the peer-reviewed literature alongside SNAP-Ed objectives. Specific behaviors and knowledge associated with the *Dietary Guidelines for Americans*[13] were chosen as fundamental indicators: consumption of fruits, vegetables, low-fat dairy, and whole grains; water and sugar-sweetened beverage intake; time spent doing physical activity and being sedentary; MyPlate knowledge; and knowledge of the Physical Activity Guidelines for Americans. Questionnaire items were generated (Stage 2) from the validated Day in the Life Questionnaire[14] and School Physical Activity and Nutrition questionnaire.[11] The resultant, visually-enhanced measure was piloted iteratively over three years for face validity (Stage 3) with fourth to eighth graders in Arizona's SNAP-Ed-eligible schools. The revised questionnaire included child-friendly font and a readability grade of 2.6 to accommodate fourth graders below reading level.

Study Design

This study tested the revised measure for internal consistency and test-retest reliability (Stage 4). The project was approved by the Arizona Department of Health Services Human Subjects

Review Board as of minimal risk to participants and administered in compliance with school district regulations for parental and student consent. A written disclaimer reinforced that participation was optional, and student assent was obtained verbally prior to questionnaire administration.

A convenience sample of 119 fourth graders aged 9 to 11 were recruited from five SNAP-Ed-eligible classrooms in Pima County, Arizona. A trained proctor administered the questionnaire in classrooms at two time points with no nutrition education in between. Four classes received the test and retest one week apart; one class was retested at six weeks to accommodate the teacher's schedule. A standardized proctor protocol was used with item explanations and prescribed responses to potential questions: the proctor led classes through each item, allowing students to ask clarifying questions but not otherwise talk or share answers. Administration times ranged from 10 to 20 minutes, depending on students' questions and completion pace.

Data Analysis

Items were categorized into primary scales and secondary subscales: Behavior—Nutrition and Physical Activity, Knowledge—Nutrition and Physical Activity. Internal consistency was examined using standardized values of Cronbach α at both time points, using reverse coding of answers for three items (refined grain consumption, sweetened beverage consumption, and sedentary time). The α values for scales and subscales were generated separately. While interpretation of α is not rigorously defined, 0.7 is generally regarded as acceptable and was used here.[8,15]

Test-retest reliability was assessed for continuous variables using the intraclass correlation coefficient (ICC) in a two-factor mixed effects model testing for consistency. Cohen's Kappa statistic, a reliability measure that takes into account the agreement occurring by chance, was

used for categorical variables. The weighted Kappa statistic was used for ordinal data. Interpretations of the ICC and Kappa were based on the values suggested by Cichetti[16] and Landis and Koch,[17] respectively. Stata/IC13.1 (StataCorp, College Station, TX, 2013) was used for all analyses.

RESULTS

Internal Consistency

Table 2 provides reliability results for internal consistency. Cronbach α was adequate[16] for the Behavior and Knowledge scales (.71 and .72, respectively) and the Nutrition Behavior and Nutrition Knowledge subscales (.78 and .75, respectively). At .46, internal consistency for the Physical Activity Behavior subscale was below the minimally acceptable level.

Test-Retest Reliability

Test-retest reliability (Table 3) was generally acceptable. ICCs for continuous data fell within the fair (0.40) to excellent (0.75) range suggested by Cichetti,[16] excluding the physical education (PE) question. Kappa coefficients for categorical and ordinal data had fair (0.30) to substantial (0.72) agreement, per Landis and Koch.[17]

DISCUSSION

The current study describes progress toward a feasible, appropriate, and valid SNAP-Ed questionnaire for school-aged youth using established best practices for developing such measures.[5,8] The KAN-Q was purposefully designed to assess school-based nutrition and physical activity interventions targeting grades 4 through 8. Its brief administration and absence of biometric data collection help to overcome the potential barriers of limited classroom time,

sociopolitical obstacles in schools, and restrictions on the time and resources of SNAP-Ed implementing agencies. The KAN-Q was also developed to align with SNAP-Ed objectives. Recently, this alignment was scrutinized against the SNAP-Ed Evaluation Framework[9] and found to measure two of its four core indicators: Healthy Eating Behaviors and Physical Activity & Reduced Sedentary Behaviors.

Beyond practicality and suitability, the KAN-Q was shown to be reliable in terms of internal consistency and reproducibility among the youngest intended audience. For internal consistency, Cronbach α exceeded .7 for the two primary scales, Behavior (α = .71) and Knowledge (α = .72), and for all subscales except Physical Activity Behavior. Hall et al.[18] reported similar findings for a nutrition-only youth survey, which included a behavior subscale (α = .71) and a knowledge subscale (α = .56). While debate exists regarding minimally acceptable α values,[15,16] Townsend[8] recommends a .6 to .7 cutoff for nutrition education measures, and researchers commonly report these values as adequate.[7,19,20] Tavakol and Dennick[15] aver that α values are positively influenced by number of scale items and negatively influenced by scale heterogeneity. Here, α values may be lower-bound estimates of reliability because the number of scale items is relatively low and multiple factors likely underlie both scales.

Test-retest reliability was acceptable for all but the PE question. The results are comparable to estimates for similar instruments.[11,19-21] In a child nutrition questionnaire, Wilson et al.[19] reported ICCs of 0.57, 0.66, and 0.66 for water, fruit, and vegetable intake, respectively, compared with KAN-Q values of 0.61, 0.75, and 0.55. Conversely, the KAN-Q had higher ICCs than the child nutrition questionnaire for fruit knowledge (0.46 vs. 0.16) and vegetable knowledge (0.51 vs. 0.36) and a lower ICC for sweetened beverage intake (0.45 vs. 0.59). In a child food frequency

questionnaire, Saeedi et al.[21] also described ICCs for fruit (0.63) and vegetable (0.60) consumption that are similar to KAN-Q findings.

With an ICC of -0.50, the PE question was likely problematic because PE classes were scheduled to recur weekly, not daily. Reframing the question from "yesterday" to ask about weekly PE may strengthen agreement. The relatively low ICC for afterschool activities (0.40) may require similar revision, as sports and activity clubs are often scheduled weekly. Additionally, four of the five Nutrition Knowledge Subscale items had acceptable but lower-range test-retest agreement: recommended cups of fruit per day, recommended cups of vegetables per day, how much of kids' plate should be covered by fruits and vegetables, and how much of all grains consumed should be whole grains. For these questions, reliability may be improved by having proctors remind students to avoid guessing and select "I don't know" when uncertain.

It is important to note that one class received pre-posts at a substantially longer time interval than the other four classes (six weeks versus one week) to accommodate the teacher's schedule, which appears to have influenced results. Agreement generally improved when the class surveyed six weeks apart was excluded from the analysis (Table 3). This suggests that six weeks may be too long an interval to test for KAN-Q agreement among this age group and that the slightly higher agreements reported in parentheses in Table 3 may be a more accurate representation of test-retest reliability.

This study had several limitations. There is an inherent challenge to developing a practical SNAP-Ed questionnaire for school-based administration: limiting length also limits scope and scale accuracy.[15] Here, internal consistency was influenced by the number of questions in each scale/subscale; dimensions with few items may have demonstrated lower-bound reliability.

Moreover, the self-report questionnaire may be less burdensome than direct observation or food records, but it is subject to recall bias.[10,11] The low reliability of the Physical Activity Behavior subscale prompted investigator doubt regarding recall accuracy. While the KAN-Q poses behavioral questions about *yesterday* to enhance recall, those items cannot be assumed to reflect *usual* intake without multiple administrations to the same cohort to better reflect habitual consumption.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Given its success as a reliable measure, the KAN-Q shows promise as a standard SNAP-Ed questionnaire for grades 4 through 8. With a 15- to 20-minute administration protocol, it addresses feasibility concerns for school-based administration, and it aligns directly with national SNAP-Ed evaluation priorities. Items that have been identified here as problematic (e.g., Physical Activity Behavior questions) are currently under revision. A potential next step is to assess convergent validity using 24-hour recall (Table 1, Stage 5).[8]

Today's SNAP-Ed initiatives demand multi-level programming using a combination of direct education with policy, systems and environmental interventions and social marketing.[1] Population indicators are sought across *all levels* for: fruits and vegetables, whole grains, dairy, beverages, physical activity and reduced sedentary behaviors.[9] Because the KAN-Q includes each of these population-level indicators, it has potential for broad use in measuring SNAP-Ed outcomes related to multiple levels of intervention, given the proper evaluation design.

REFERENCES

1. USDA, Food and Nutrition Service. SNAP-Ed plan guidance FY 2017: nutrition education and obesity prevention grant program. https://snaped.fns.usda.gov/sites/default/files/uploads/Final%20FY%202017%20SNAP-Ed%20Plan%20Guidance%2004-04-2016.pdf. Accessed July 30, 2016.

2. Institute of Medicine. *Nutrition Education in the K-12 Curriculum: The Role of National Standards: Workshop Summary*. Washington, DC: The National Academies Press; 2013.

3. Institute of Medicine. *Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation.* Washington, DC: The National Academies Press; 2012.

4. Budd GM, and Volpe SL. School-Based Obesity Prevention, Research, Challenges, and Recommendations. *J Sch Health*. 2006;76(10):485-495.

5. Guthrie JF, Stommes E, Voichick J. Evaluating Food Stamp Nutrition Education: issues and opportunities. *J Nutr Educ Behav*. 2006;38:6-11.

6. Taylor-Powell E. Evaluating Food Stamp Nutrition Education: a view from the field of program evaluation. *J Nutr Educ Behav*. 2006;38:12-17.

7. Branscum P, Sharma M, Kaye G, and Succop P. An evaluation of the validity and reliability of a food behavior checklist modified for children. *J Nutr Educ Behav*. 2010;42:349-352.

8. Townsend MS. Evaluating Food Stamp Nutrition Education: Process for development and validation of evaluation measures. *J Nutr Educ Behav*. 2006;38(1):18-24.

9. USDA, Food and Nutrition Service. The SNAP-Ed Evaluation Framework: Nutrition, Physical Activity, and Obesity Prevention Indicators: Interpretive Guide to the SNAP-Ed Evaluation Framework. 2016. https://snaped.fns.usda.gov/. Accessed July 30, 2016.

10. Thiagarajah K, Fly A, Hoelscher D, Bai Y, Lo K, Leone A, and Shertzer J. Validating the food behavior questions from the elementary school SPAN Questionnaire. *J Nutr Educ Behav*. 2008;40:305-310.

11. Penkilo M, George GC, Hoelscher DM. Reproducibility of the School-based Nutrition Monitoring Questionnaire among Fourth-grade Students in Texas. *J Nutr Educ Behav*. 2008;40(1):20-27.

12. Neuhouser ML, Lilley S, Lund A, Johnson DB. Development and Validation of a Beverage and Snack Questionnaire for Use in Evaluation of School Nutrition Policies. *J Am Diet Assoc.* 2009;109(9):1587-1592.

13. US Department of Health and Human Services and USDA. 2015–2020 Dietary Guidelines for Americans. 8th Edition. December 2015. http://health.gov/dietaryguidelines/2015/guidelines/. Accessed July 30, 2016.

14. Edmunds L, Ziebland S. Development and validation of the Day in the Life Questionnaire (DILQ) as a measure of fruit and vegetable questionnaire for 7–9 year olds. *Health Educ Res.* 2002;17(2):420–430.

15. Tavakol M and Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-55.

16. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6(4):284–290.

17. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

18. Hall E, Chai W, Koszewski W, and Albrecht J. Development and validation of a social cognitive theory-based survey for elementary nutrition education program. *Int J of Behav Nutr Phys Act.* 2015;12:47.

19. Wilson AM, Magarey AM, Mastersson N. Reliability and relative validity of a child nutrition questionnaire to simultaneously assess dietary patterns associated with positive energy balance and food behaviours, attitudes, knowledge and environments associated with healthy eating. *Int J Behav Nutr Phys Act.* 2008;5(1):5.

20. Gray HL, Koch PA, Contento IR, Bandelli LN, Ang I, Di Noia J. Validity and reliability of behavior and theory-based psychosocial determinants measures, using audience response system technology in urban upper-elementary schoolchildren. *J Nutr Educ Behav.* 2016; 48:437-452.

21. Saeedi P, Skeaff SA, Wong JE, Skidmore PML. Reproducibility and relative validity of a short food frequency questionnaire in 9-10 year-old children. *Nutrients.* 2016;8:271.

Table 1. Best Practice Development of Nutrition Education Measures, and Progress-to-date for the Kids' Activity and Nutrition Questionnaire (KAN-Q)

Stage of Development	KAN-Q Progress
1: *Domain selection* using literature review	Completed
2: *Item generation* from evidence base and expert contribution	Completed
3: *Item pre-testing* with target audience	Completed
4: *Item testing/analysis* for consistency and reliability	Reported here
5: *Convergent and criterion validity testing*	Planning phase
6: *Sensitivity assessment* of ability to detect change	Planning phase

Table 2. Internal Consistency Reliability for Two Administrations of the Kids' Activity and Nutrition Questionnaire (KAN-Q)

Scale/Subscale	No. of Questions	Cronbach α (test 1)	Cronbach α (test 2)
Behavior	13	.72	.71
Nutrition	7	.79	.78
Physical Activity	6	.36	.46
Knowledge	5	.52	.72
Nutrition	4	.58	.75
Physical Activity	1	NA	NA
Overall	18	.73	.73

NA, not available

Table 3. Test-Retest Reliability for the Kids' Activity and Nutrition Questionnaire (KAN-Q)

Scale/Subscale	Item	ICC[a]
Behavior		
Nutrition	Yesterday, did you drink any milk?[b]	0.68 (0.70)
	Yesterday, did you eat any corn tortillas or bread, tortillas, buns, bagels or rolls that were *brown*?[b]	0.43 (0.50)
	Yesterday, did you eat any corn tortillas or bread, tortillas, buns, bagels or rolls that were *white*?[b]	0.57 (0.59)
	Did you eat any vegetables yesterday?[b]	0.55 (0.54)
	Yesterday, did you eat any fruit?[b]	0.75 (0.76)
	Yesterday, did you drink any regular (*not* diet) soda, Kool-Aid®, sports drink, or other fruit-flavored drinks?[b]	0.45 (0.45)
	Yesterday, did you drink any water?[b]	0.61 (0.61)
Physical Activity	Did you do any activities after school yesterday that made your heart beat fast or made you breathe hard?[c]	0.40 (0.45)
	Did you go to physical education (PE) or gym class yesterday?[c]	-0.50 (-0.61)
	When you were not in school yesterday, how many hours did you spend sitting or lying around?[c]	0.48 (0.53)
Knowledge		
Nutrition	How many total cups of fruit should most kids eat each day?	0.46 (0.57)
	How many total cups of vegetables should most kids eat each day?	0.51 (0.54)
Physical Activity	How many minutes of physical activity/exercise should most kids have on all or most days of the week?	0.53 (0.62)
		Kappa
Behavior		
Nutrition	What type of milk do you drink most of the time?	0.72 (0.74)[d]
Physical Activity	How did you travel *to* school yesterday?	0.56 (0.61)[d]
	How did you travel *home from* school yesterday?	0.53 (0.57)[d]
	What did you do most of the time at lunchtime recess yesterday?	0.38 (0.38)[e]
	If you had a morning break yesterday, what did you do most of the time at morning break?	0.55 (0.60)[e]
Knowledge		

Nutrition	What type of milk should most kids drink most of the time?	0.48 (0.59)[d]
	How much of most kids' plates at meals should be fruits and vegetables?	0.39 (0.46)[e]
	How much of the bread and cereal that most kids eat should be made with whole grains (brown, whole wheat, etc.)?	0.30 (0.42)[e]

ICC, intra-class correlation coefficient; [a]Values that exclude the class surveyed at a six-week (versus one-week) time interval are in parentheses; [b]Response choices were continuous as number of times consumed; [c]Response choices were continuous as number of minutes/hours; [d]Cohen's Kappa statistic; [e]Weighted kappa statistic