



Published in final edited form as:

*Am J Health Promot.* 2011 July ; 25(6): 417–421. doi:10.4278/ajhp.090914-ARB-297.

## Psychometric Characteristics of Process Evaluation Measures for a Rural School-based Childhood Obesity Prevention Study: Louisiana Health

R. L. Newton Jr., J. L. Thomson, K. Rau, S. Duhe', A. Sample, N. Singleton, S. D. Anton, L. S. Webber, and D. A. Williamson

### Abstract

**Purpose**—To evaluate the implementation of intervention components of the Louisiana Health study, which was a multi-component childhood obesity prevention program conducted in rural schools.

**Design**—Content analysis.

**Setting**—Process evaluation assessed implementation in the classrooms, gym classes, and cafeterias.

**Subjects**—Classroom teachers (n = 232), physical education teachers (n = 53), food service managers (n = 33), and trained observers (n = 9).

**Measures**—Five process evaluation measures were created: Physical Education Questionnaire (PEQ), Intervention Questionnaire (IQ), Food Service Manager Questionnaire (FSMQ), Classroom Observation (CO) and School Nutrition Environment Observation (SNEO).

**Analysis**—Inter-rater reliability and internal consistency were conducted on all measures. ANOVA and Chi-square were used to compare differences across study groups on questionnaires and observations.

**Results**—The PEQ and one sub-scale from the FSMQ were eliminated because their reliability coefficients fell below acceptable standards. The sub-scale internal consistencies for the IQ, FSMQ, CO, and SNEO (all Cronbach's  $\alpha > .60$ ) were acceptable.

**Conclusions**—After the initial 4 months of intervention, there was evidence that the Louisiana Health intervention was being implemented as it was designed. In summary, four process evaluation measures were found to be sufficiently reliable and valid for assessing the delivery of various aspects of a school-based obesity prevention program. These process measures could be modified to evaluate the delivery of other similar school-based interventions.

### Keywords

process evaluation; internal validity; psychometrics; weight gain prevention intervention

### Purpose

The Louisiana Health study was a large scale, multi-component, school-based health promotion program designed to test the efficacy of two school-based childhood obesity prevention interventions. It is one of several school-based obesity prevention programs. The results of these school-based programs have varied in terms of outcomes, with some, but not all, reporting weight changes.<sup>1</sup> Process evaluations serve a number of functions for an intervention, including monitoring its progress, assessing implementation, and explaining the results.<sup>2–4</sup> However, few school-based obesity prevention studies have conducted a

process evaluation.<sup>5</sup> The main purpose of the current manuscript is to describe the development of the process evaluation measures for the Louisiana Health study, which was conducted in rural Louisiana schools performing below state academic levels. The process evaluation was specifically designed to assess the implementation and quality control of the three study groups in the Louisiana Health study. A secondary purpose is to describe the fidelity of the intervention after the initial implementation.

## Methods

### Design

The 28-month study compared the effectiveness of a dietary environmental intervention and a dietary + Internet behavioral intervention to a no-treatment control group to prevent excess weight gain in 4<sup>th</sup> – 6<sup>th</sup> grade students. The study enrolled 2101 students in 17 school clusters. There were 10 schools in the dietary environmental intervention, 14 in the dietary + Internet behavioral intervention, and 9 in the control group. A more in-depth description of the Louisiana Health study and its dietary environmental (previously <sup>6</sup> labeled, Primary) and dietary + Internet behavioral (previously <sup>6</sup> labeled, Primary + Secondary) programs can be found elsewhere.<sup>6</sup> Informed consent was obtained from the parents, and assent was given by the students who participated in the study. The study was reviewed and approved by the Institutional Review Board of the Pennington Biomedical Research Center.

### Sample

The classroom teachers (n = 232), physical education teachers (n = 53), and food service managers (n = 33) completed the questionnaires, and trained observers (n = 9) conducted observations of the classrooms and cafeterias. Specifically, the classroom teachers completed the Intervention Questionnaire, the physical education (PE) teachers completed the Physical Education Questionnaire, and food service managers completed the Food Service Manager questionnaire. Trained observers conducted the Classroom Observation and School Nutrition Environment Observation (observation of the cafeterias). The trained observers were Louisiana Health research staff members. These individuals received detailed instructions on how to conduct the observations, including understanding the purpose of the evaluations, where to locate specific items in the classrooms and cafeterias, and how to be unobtrusive. The observers were trained by more senior Louisiana Health staff members that had previous experience conducting process evaluations. All teachers and food service managers were assessed, and all cafeterias at each school were observed. However, only one classroom per grade was randomly observed due to logistical constraints. Teachers and cafeteria staff were not informed when observations would take place.

### Measures

All process measures were created after the baseline assessment of participants, which occurred from August through December of 2006. Each measure was initially created by the primary author and then a Q-sort <sup>7</sup> of these initial items was conducted by eight Louisiana Health research staff members. The Q-sort is a method of item selection, in which a respondent is given each item on a scale and sorts them into piles of pre-specified categories. The Q-sort resulted in the elimination of items from several measures, and these item-reduced instruments were subsequently used in the process evaluation that was conducted in April and May of 2007. The reliability and validity of the instruments were assessed after the process evaluation data were collected.

Each process evaluation instrument included items that were relevant and representative of program components assessed. The questionnaires were designed to gather self-reported data from teachers and food service managers on program implementation. The observations

were created to provide objective assessments of program implementation in the classrooms and cafeterias. All instruments were scored by summing the items within a sub-scale.

**Physical Education Questionnaire (PEQ)**—The scale assessed the activities occurring in PE class. The dietary environmental sub-scale assessed for the implementation of the Sports, Play and Active Recreation for Kids (SPARK; San Diego, CA ©) curriculum. The control items pertained to typical PE activities. The 6-item Likert-type scale was scored 0 (Strongly Disagree) to 4 (Strongly Agree).

**Intervention Questionnaire (IQ)**—The scale assessed for activities in each study group that should be occurring according to the study design. There were dietary environmental, dietary + Internet behavioral, control and general (commonly used teaching strategies) sub-scales. The 28-item Likert-type scale was scored 0 (Strongly Disagree) to 4 (Strongly Agree).

**Food Service Manager Questionnaire (FSMQ)**—This scale assessed for food preparation styles in the cafeteria. The recommended sub-scale assessed for Louisiana Health suggested alterations and the non-recommended sub-scale assessed for common, but unhealthy practices. The 24-item Likert-type scale was scored 0 (Strongly Disagree) to 4 (Strongly Agree).

**Classroom Observation (CO)**—The CO measured the presence or absence of Louisiana Health recommended alterations to the classroom. The dietary environmental sub-scale assessed for Louisiana Health environmental changes, and the control sub-scale pertained to elements of typical classroom environments. The 16-item scale was scored Yes (1) or No (0).

**School Nutrition Environment Observation (SNEO)**—The SNEO assessed the cafeteria environment. The recommended sub-scale was composed of items related to the suggested alterations to the cafeteria environment. The non-recommended sub-scale assessed for unhealthy elements. The 56-item scale was scored Yes (1) or No (0).

## Analysis

Items from each scale were eliminated if there was less than 75% agreement on the Q-sort. The subsequent reliability testing was performed on the Q-sort item reduced scales. Inter-rater agreement on the Q-sort was assessed using Gwet's AC1 statistic.<sup>8</sup> Internal consistency was assessed using Cronbach's  $\alpha$  coefficients. The instruments were considered to be adequately reliable if the overall reliability was  $\geq .60$ .<sup>9;10</sup> Fidelity of the IQ was assessed through analysis of variance (ANOVA) and Kruskal-Wallis ANOVA ( $\chi^2$ ) was used for the other measures. All statistical analyses were performed using SAS® software, version 9.1 or 9.2 (SAS Institute Inc., Cary, NC) and SPSS® software, version 16.0 (SPSS Inc, Chicago, IL). Results were considered significant at an alpha level of  $p < .05$ .

## Results

Table 1 summarizes the psychometric data from the process evaluation.

**Physical Education Questionnaire (PEQ)**—The Q-sort did not result in any item being eliminated from the PEQ, which consisted of 3 dietary environmental and 3 general items. The inter-rater reliability (AC1 = .25) did not meet the reliability criteria. Therefore, it

was deemed too unreliable to be used in the study and no further analyses were conducted utilizing this scale.

**Intervention Questionnaire (IQ)**—Based upon the results of the Q-sort, 4 items were eliminated from the IQ, which left 6 dietary environmental, 9 dietary + Internet behavioral, 5 control, and 4 general items. The inter-rater reliability ( $AC1 = .75$ ) and internal consistencies ( $\alpha = .77 - .85$ ) for all sub-scales met the criterion for satisfactory reliability.

**Food Service Manager Questionnaire (FSMQ)**—Only 1 item was eliminated from the FSMQ after the Q-sort, leaving 13 dietary environmental and 10 general items. The inter-rater reliability ( $AC1 = .94$ ) and internal consistency ( $\alpha = .76$ ) for the recommended sub-scale met the criterion for satisfactory reliability.

**Classroom Observation (CO)**—The Q-sort and Cronbach alpha resulted in 9 items being removed, leaving 3 dietary environmental and 4 control items. The inter-rater reliability ( $AC1 = .89$ ) and internal consistencies for both sub-scales ( $\alpha = .63$  &  $.60$ ) met the criterion for satisfactory reliability.

**School Nutrition Environment Observation (SNEO)**—There were 6 items excluded based upon low agreement from the Q-sort and 25 additional items were eliminated based on extremely low response rates ( $< 50\%$ ). The final scale therefore consisted of 23 recommended and 2 non-recommended items. The inter-rater reliability ( $AC1 = .73$ ) and internal consistencies for both sub-scales ( $\alpha = .77 - .85$ ) met the criterion for satisfactory reliability.

**Fidelity**—It was expected that teachers/food service managers in their respective study groups would score higher than teachers who were not in that study condition on all questionnaires. Teachers delivering the dietary environmental intervention had higher mean scores ( $F [2, 176] = 45.1, p < .001$ ) on the dietary environmental sub-scale of the IQ compared to control teachers (Figure 1). Teachers receiving the dietary + Internet behavioral intervention scored higher ( $F [2, 178] = 40.5, p < .001$ ) on the dietary + Internet behavioral sub-scale compared to teachers not receiving the intervention. There were no differences between the groups on sub-scales assessing implementation of control intervention or general teaching practices. For the FSMQ, no significant intervention group mean differences were found.

It was expected that scores would be higher on dietary environmental/recommended sub-scales for intervention classrooms/cafeterias compared to control groups on all observations, with no differences on control/non-recommended sub-scales. For the CO, intervention classrooms received significantly higher mean scores than control classrooms ( $\chi^2 = 17.7, p < .001$ ) on the dietary environmental sub-scale. For the SNEO, intervention cafeterias received significantly higher mean scores on the recommended sub-scale than cafeterias in the control group ( $\chi^2 = 10.6, p = .005$ ). Although there was overall significance ( $\chi^2 = 6.1, p = .048$ ) for the presence of differences in the non-recommended sub-scale, no significant pair-wise differences were found (Figure 1).

## Discussion

### Summary

This paper describes the first attempt to establish process measures for prevention programs delivered in rural schools performing below state academic standards.<sup>6</sup> Five measures were developed to assess the internal validity of the Louisiana Health study. The reliability of

each measure was assessed through inter-rater reliability and internal consistency. One entire measure (PEQ) and the sub-scale from one other (FSMQ) were eliminated because their reliability coefficients fell below acceptable standards. After elimination of the PEQ, four reliable measures (IQ, FSMQ, CO, SNEO) containing nine sub-scales were tested. In order to determine if the study was being implemented according to the study design, study group differences on the four measures were assessed. The significant study group differences for the IQ, CO, and SNEO in the expected direction suggested that Louisiana Health had evidence of program fidelity. Despite fidelity in terms of the classroom and the cafeteria environment, one area of improvement lies with the implementation by the food service managers. These managers reported similar cooking styles regardless of the intervention condition.

### Limitations

The small sample size for cafeteria measures was a limitation of the study. This was due to the fact that there were only a limited number of food service managers and cafeterias eligible to be assessed, and missing data lowered this number. In addition, we assumed the environmental nature of the program would allow it to reach all children. Although this is likely because the intervention took place daily in the cafeteria and classrooms, there are no data supporting this assumption.

### Significance

Our study is significant because it reports on the development of process evaluation methods for school-based obesity prevention programs that were implemented in rural schools performing below state academic standards. Although many school-based programs have been conducted,<sup>1</sup> few have performed a process evaluation.<sup>11</sup> Our approach to process evaluation included conducting observations, administering questionnaires, developing measures that could be administered across study groups, and assessing reliability. These methods can be utilized in developing future process evaluation measures and expanded to assess other areas of process evaluation, such as dose and reach.<sup>4</sup>

As stated above, school-based obesity prevention programs have infrequently resulted in weight changes between active prevention programs and control groups.<sup>11</sup> These findings may be due to lack of proper implementation of the interventions, in terms of dose, fidelity, or reach.<sup>5</sup> The process measures that were developed in this study will be used to help explain the outcomes of the Louisiana Health study. We have shown that, at least initially, the program is largely being conducted according to intervention design. If we continue to see differences in the expected direction, it will provide evidence of program integrity and would increase the likelihood that any effects are a direct result of the intervention.<sup>3</sup>

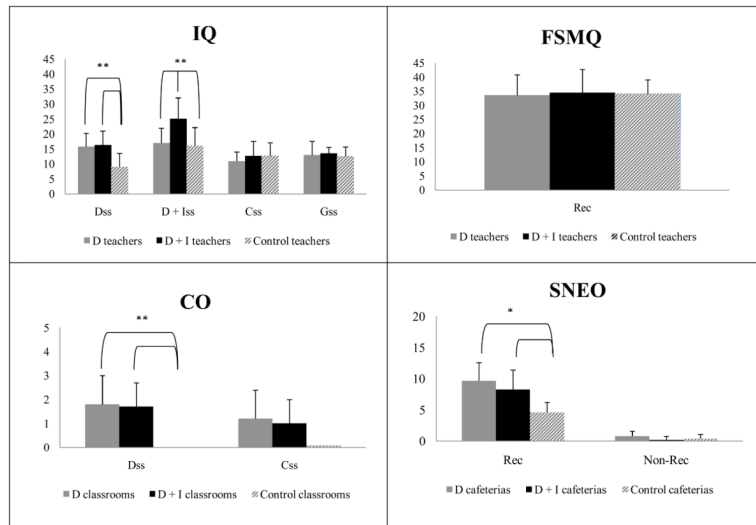
### Acknowledgments

This work was supported by the National Institute for Child Health and Human Development of the National Institutes of Health [R01 HD048483]; and the United States Department of Agriculture [58-6435-4-90]; and was partially supported by the NORC Center Grant #1P30 DK072476 entitled Nutritional Programming: Environmental and Molecular Interactions sponsored by NIDDK. The clinical trial number for this project is: NCT00289315.

### References

1. Sharma M. School-based interventions for childhood and adolescent obesity. *Obes Rev.* 2006; 7:261–269. [PubMed: 16866974]
2. McGraw SA, Stone EJ, Osganian SK, et al. Design of process evaluation within the Child and Adolescent Trial for Cardiovascular Health (CATCH). *Health Educ Q.* 1994; Suppl 2:S5–S26. [PubMed: 8113062]

3. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. *Health Promot Pract.* 2005; 6:134–147. [PubMed: 15855283]
4. Steckler, A.; Linnan, L. *Process Evaluation for Public Health Interventions and Research.* San Francisco, CA: Jossey-Bass; 2002.
5. Kropski JA, Keckley PH, Jensen GL. School-based obesity prevention programs: an evidence-based review. *Obesity (Silver Spring).* 2008; 16:1009–1018. [PubMed: 18356849]
6. Williamson DA, Champagne CM, Harsha D, et al. Louisiana (LA) Health: design and methods for a childhood obesity prevention program in rural schools. *Contemp Clin Trials.* 2008; 29:783–795. [PubMed: 18448393]
7. Stephenson W. The Study of Behavior: Q-Technique and Its Methodology. *Psychometrika.* 1954; 19:327–333.
8. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol.* 2008; 61:29–48. [PubMed: 18482474]
9. Schmitt N. Uses and abuses of coefficient alpha. *Psychol Assess.* 1996; 8:350–353.
10. Puhan MA, Bryant D, Guyatt GH, Heels-Ansdell D, Schunemann HJ. Internal consistency reliability is a poor predictor of responsiveness. *Health Qual Life Outcomes.* 2005; 3:33. [PubMed: 15877824]
11. Thomas H. Obesity prevention programs for children and youth: why are their results so modest? *Health Educ Res.* 2006; 21:783–795. [PubMed: 17099075]



**Figure 1.** Differences in mean scores on process measures. Error bars represent standard deviation. \* $p < .005$ , \*\* $p < .001$ .

Note. Dss = Dietary environmental sub-scale; D + Iss = Dietary + Internet behavioral sub-scale; Ccss = Control sub-scale; Gss = General sub-scale; Rec = Recommended sub-scale; Non-rec = Non-recommended sub-scale; D = Dietary environmental; D + I = Dietary + Internet behavioral.

**Table 1**  
Inter-rater reliability and internal consistency for process measures in the Louisiana Health study.\*

Process Evaluation Measure	Number of Items	Inter-rater reliability			Internal Consistency	
		AC1	SE	p	n	Cronbach's alpha
PE Questionnaire (PEQ)	6	.25	.09	0.005		
Intervention Questionnaire (IQ)	24	.75	.03	< 0.001	174	
Dietary environmental	6					.85
Dietary + Internet behavioral	9					.86
Control	5					.79
General	4					.77
Food Service Manager Questionnaire (FSMQ)	23	.94	.05	< 0.001	26	
Recommended	13					.76
Non-recommended	10					.37
Classroom Observation (CO)	7	.89	.10	< 0.001	65	
Dietary environmental	3					.63
Control	4					.60
School Nutritional Environment Observation (SNEO)	14	.89	.03	< 0.001	26	
Recommended	12					.73
Non-recommended	2					.63

\* AC1 indicates chance-adjusted agreement statistic; SE, standard error. The SE was based upon unconditional variance.