Evaluating the Wellness School Assessment Tool for Use in Public Health Practice to Improve School Nutrition and Physical Education Policies in New York

IAN BRISSETTE, PhD a KATHLEEN WALES, BS b MEGHAN O’CONNELL, MPH c

ABSTRACT

BACKGROUND: Addressing the limitations of existing Local Wellness Policies (LWPs) and promoting their implementation remain priorities for health and education agencies. One gap has been the absence of a standard assessment to support LWP revision. During planning for an initiative to improve school nutrition and physical education policy, the Wellness School Assessment Tool (WellSAT) was evaluated.

METHODS: Five public health practitioners used WellSAT to assess 50 LWPs. A randomized, counterbalanced design ensured each LWP was coded twice by separate raters. Models evaluated the extent to which WellSAT ratings reflected differences in the LWPs and order, familiarity and rater effects. During field testing, 18 public health practitioners used WellSAT as part of a statewide public health initiative.

RESULTS: In pilot testing the majority of the variability in WellSAT scores (median = 88%; range = 76% to 100%) was attributable to differences between policies. Correlations between independent raters’ strength and comprehensiveness scores were strong, r = .88 and r = .77, respectively. During field testing, WellSAT was well accepted by public health practitioners and members of the school community.

CONCLUSIONS: WellSAT represents a reliable and feasible tool for health and education agencies to use in improving LWPs and aligning them with recognized standards.

Keywords: evaluation; health policy; public health.


Establishing school wellness policies with strong, specific language has been recommended as a means to foster school environments that encourage healthy eating and lifelong physical activity (PA). 1 2 The Local Wellness Policy (LWP) requirements included in the Child Nutrition and Women, Infants, and Children (WIC) Reauthorization Act of 2004 (Public Law 108-265) provided a means to improve school nutrition and increase quality physical education (PE) and PA for millions of children in the United States. State and national assessments following the Child Nutrition and WIC Reauthorization Act indicated a high rate of compliance with the LWP mandate and suggested the quality of the LWPs improved during the 3 years following its enactment. 3-5 However, gaps between the content of written LWPs and school wellness standards remain. For example, few LWPs include policies that align with National Alliance for Sport and...
Physical Activity best practice recommendations about required amounts of PE or Institute of Medicine recommendations about the nutrient content of food sold in schools (eg, Institute of Medicine). More specifically, differences between what is written in the LWPs and what is practiced are assumed to exist due to lack of technical assistance and resources to support implementation. Although the 2010 Child Nutrition and WIC Reauthorization law (titled The Healthy and Hunger-free Kids Act) includes additional provisions to improve the school nutrition environment, addressing the limitations of existing LWPs and supporting their implementation remain priorities for health and education agencies.

For over a decade, the New York State Department of Health (NYSDOH) has funded programs providing technical support and resources to assist schools with developing and implementing policies that improve the nutritional value of foods available, and increase access to quality PE and PA. This approach has used school wellness policies as a means to promote health among students, faculty, and staff. Regional consultants engaged members of the school community in the completion of a self-assessment of their policies and practices around nutrition, PE and PA, and worked with the community to develop and implement policies addressing identified priorities. Since 2004, this approach has used the policy language in the required LWPs as the vehicle for instituting environment and practice changes within schools. Consultants convened local wellness committees to evaluate the contents of LWPs, facilitated revisions to the LWPs to align the written policies with standards and existing state regulations, and provided technical assistance to local wellness committees, teachers, and staff on policy implementation.

One gap had been the lack of a standard tool for assessing existing policy language contained within the LWPs. For the regional consultants, having a standard assessment could better enable them to illustrate the strengths and deficiencies of policies within an LWP to the local wellness committees, school board, school administration and staff, parents and other stakeholders. For the school community, the information from a standard policy assessment tool could be used as part of developing improvement plans and assessing progress toward policy goals. For the NYSDOH, a standard assessment could yield consistent and reliable information about the LWPs of school districts reached. Existing research tools provided the rigor required to quantify the strengths and weaknesses of LWPs, but the length of time involved in completing these assessments and technical skills required to tabulate and summarize data from them precluded their use by those without research experience.

In 2009, NYSDOH began planning for a comprehensive statewide initiative, Healthy Schools NY, to expand ongoing efforts to use regional consultants and changes to LWPs as the basis for improving the school nutrition, PA, and PE environment. The Rudd Center for Food Policy and Obesity’s Wellness School Assessment Tool (WellSAT) was identified as a resource that could be used to address the previous gap of not having a standardized tool for assessing the contents of LWPs. The tool was based on a research tool used in multiple studies of the impact of LWPs, but designed to be used by state departments of public health and education and local school district level personnel. With WellSAT, regional contractors could be expected to rate the LWP of the school districts with which they were working and present the results to local wellness committees and school administration to gain support for improving the specificity, comprehensiveness, and strength of the policy language. Prior to adopting WellSAT for use in Healthy Schools NY, pilot, and field testing was completed. The purpose was to determine whether WellSAT was feasible for use by nonresearchers and to determine if public health practitioners could use it to objectively and reliably assess LWPs which they had a stake in developing. This article describes these tests and how they informed ongoing use of the WellSAT in a statewide initiative to improve nutrition and the quality and quantity of PE in schools.

METHODS

Participants
The participants were public health practitioners from 5 organizations contracted by the NYSDOH to work with schools to increase opportunities for PA and improve the nutritional quality of food available to students; 2 county health departments in NYS; a medical school associated with a large public university; a statewide nonprofit organization for the promotion of quality PE, PA, and recreational dance; and a regional nonprofit organization involved in coordinating health planning and education for a multi-county area. These organizations were recruited because of their familiarity with LWPs and because they represent traditional partners for schools and the types of organizations likely to participate in Healthy Schools NY.

Instruments and Training
The WellSAT is a standardized assessment tool for rating the strength and comprehensiveness of elements of LWP required by the Child Nutrition and WIC Reauthorization Act of 2004. WellSAT provides an online system to evaluate school wellness policies that automatically compiles data for the user. It can be completed multiple times to enable users to track progress within a district or compare across districts. The tool contains 50 items, each with examples of language from existing school wellness policies to assist...
users with scoring. Users are given a scorecard upon completion that links them to resources to improve specific parts of their school wellness policies.

WellSAT users receive school wellness policy scores in 2 domains—comprehensiveness and strength. It provides a score from 1 to 100 for the comprehensiveness and strength of the school wellness policy overall, as well as comprehensiveness and strength scores for the following subscales: nutrition education and wellness promotion; nutrition standards for the United States Department of Agriculture (USDA) meal programs; nutrition standards for competitive foods, PE, and PA; and evaluation. Each of the items were scored on a scale from 0 to 2, where 0 represented no mention of the item in the wellness policy, 1 represented stipulation of the item in weak or vague language (eg, “Vending machines should include items which are healthful”), and 2 indicated a strong and specific policy (eg, “All items sold through vending machines shall contain no more than 1 serving per package, no more than 35% of calories from sugar, and no trans fat”).

Procedures

A 45-minute conference call was held with participants to introduce WellSAT and the project. The call covered the rationale and development of the instrument and a brief demonstration of the functionality of the online assessment tool. The system demonstration included an overview of the coding instructions, procedures for data entry and data export function. No additional training on the WellSAT or rating guidance was provided as part of the pilot.

Public health practitioners from each of the 5 organizations provided LWPs from 10 school districts and sent electronic copies to the NYSDOH. Participants were instructed to select familiar LWPs to assess the effect of familiarity on ratings assigned while completing WellSAT. Identifiable information was removed from the LWPs and each participant was sent a group of 20 LWPs in a pre-determined randomized order that included the 10 LWPs they submitted and 10 LWPs the remaining four participants submitted. The assignment of policies was counterbalanced and randomized such that each LWP was coded exactly twice by separate participants. Participants recorded the time it took to code each LWP to provide information about the feasibility of WellSAT and provided written comments about individual items or sections of the scale. After rating all 20 policies using WellSAT, participants exported their data from the WellSAT website and sent the resulting flat file by e-mail. Coding occurred between April and June 2010.

Data Analysis

Data from the pilot were combined, processed, and analyzed using SAS software. A sequence of models using Proc Mixed in SAS software was used to evaluate factors contributing to variation in the WellSAT scores. Mixed models were selected to properly account for covariance associated with having participants rate multiple policies. To examine the amount of variability due to differences in policies as opposed to different raters, we computed unconditional means models where WellSAT strength and comprehensiveness scores for the entire scale and each of the 5 subscales served as the dependent variable. These models expressed WellSAT scores as a fixed effect (μij) and 2 variance components, one representing variation between raters (τij) and a second representing variation among policies rated (σ2). Results of these models were used to calculate interclass correlations (ICC (ρ)) reflecting the proportion of the total variation in WellSAT scores attributable to differences in policies. To determine whether the WellSAT was robust to rater, order effect and familiarity effects, terms representing the 5 participants (rater), the order in which policies were rated (order) and whether the practitioner rating the policy had submitted the policy for the pilot or not (familiarity) were added.

Information about the time spent coding each LWP was entered into a spreadsheet and merged with the WellSAT data. Unconditional means and mixed models were also used to examine factors contributing to variance in the time it took to code policies. A 4-level categorical variable was developed from a continuous variable representing the order in which the policies were rated to test whether the time it took to complete the WellSAT changed with increasing experience.

Field Testing Procedures

Field testing occurred when WellSAT was adopted as a mandatory assessment tool for Healthy Schools NY. Public health practitioners from 18 organizations funded by the NYSDOH Healthy Schools NY initiative participated. Relative to the pilot, the amount of training on WellSAT was increased and standardized. All 18 Healthy Schools NY contractors were required to use the WellSAT to code a single policy and submit their results. A conference call was held to discuss the coding exercise results and illustrate how data analysis could identify potential inappropriate use of the rating scales. During spring 2011 Healthy Schools NY contractors began using WellSAT as part of their ongoing work to engage school districts to initiate policy change around nutrition, PE, and PA.

RESULTS

Pilot Testing

Participants spent an average of 40 minutes rating the LWPs in the pilot using WellSAT (range = 10 to 133). In multivariate models neither familiarity with the policy (ρ = 2.13, t = .41, n.s.) nor the order...
in which a LWP was coded ($\gamma = -3.67$, $t = 1.56$, $p = .12$) were significantly associated with time spent coding. There was a single rater effect such that the 2 participants that reported the longest and shortest times differed significantly, $\gamma = 23.6$, $t = 2.79$, $p < .01$. No other differences in time spent rating the LWPs were significant.

For both the overall WellSAT scores and the scores on each of the 5 subscales, unconditional means models indicated the majority of the variance in ratings was due to differences in policies as opposed to differences between raters. As summarized in Table 1, the interclass correlation ranged from .76 to 1.0 with a median of .88. Correlations were calculated between the strength and comprehensiveness scores the independent raters assigned to policies using WellSAT. The correlation coefficients for the overall comprehensiveness and strength scores were $r = .88$, $p < .001$ and $r = .77$, $p < .001$, respectively, and the coefficients for the subscales ranged from .61 to .88, with all $p$-values < .001. Multivariate models were used to determine the contribution of order, rater, and familiarity effects to comprehensiveness and strength scores derived from WellSAT. Table 2 summarizes results of these models predicting the strength and comprehensiveness scores for the total WellSAT scale to illustrate overall findings. None of the models found statistical evidence of familiarity effect or order effect on WellSAT ratings. There was a single consistent rater effect such that a single participant consistently had statistically higher WellSAT scores on overall strength and comprehensiveness scores on 7 of the 10 subscales.

Field Testing

During field testing, it took some members of the Healthy Schools NY cohort much longer (up to 2 hours) to code the sample LWP using WellSAT than the 40-minute average recorded in the pilot. However, these individuals reported that coding subsequent LWPs required much less time. Throughout field testing members of the Healthy Schools NY cohort were able to successfully download WellSAT data files and submit them as part of routine performance monitoring. In contrast to expectations, several Healthy Schools NY contractors indicated that school district wellness committees with whom they were working expressed interest in completing the WellSAT as a group with printed copies of the rating guidance and items. Because WellSAT was incorporated into this public health initiative to support policy improvement and program evaluation, the unconventional use of the assessment was allowed.

**DISCUSSION**

Both the pilot and field test suggested WellSAT was feasible for users not accustomed to research. In the pilot, with relatively minimal training, all 5 public health practitioners were able to use the online application successfully to code their assigned policies, download data files, and submit them through e-mail. The length of the policies was the major determinant of the time spent coding and there was no evidence of pilot participants taking significantly longer to use WellSAT to rate the initial LWPs. As compared to participants in the pilot, those who participate in the field test reported a steeper learning curve completing WellSAT. However, all 18 Healthy Schools NY contractors were able to use WellSAT to assess the LWPs of school districts, share the results with local wellness committees and download and submit their data. Taken together the results provide support for WellSAT being a practical tool for education or health agencies to collect information from multiple users in a standard format.

With regard to the reliability of WellSAT, the pilot found that nearly 90% of the variance in the strength and comprehensiveness scores assigned was due to variance between policies, suggesting the differences in the ratings assigned reflect true differences in the language included in the LWPs. There was no evidence of participants showing bias when rating policies with which they were familiar. Moreover, the WellSAT scores raters assigned did not change with repeated use. Collectively these results provide strong support for WellSAT being a reliable assessment to use in public health practice to improve and evaluate improvements to LWPs.

| Table 1. Estimated Variance in WellSAT Scores Attributable to Differences Between Coders and Policies |
| Variance between policies, $\sigma^2$ | Variance between coders, $\tau_{00}$ | Interclass correlation, $\rho$ |
| Total WellSAT | 292.9 | 51.1 | .85 |
| | 396.6 | 60.8 | .87 |
| Nutrition Education and Wellness Promotion | 457.0 | 53.5 | .89 |
| | 454.6 | 0 | 1.0 |
| USDA Standards | 758.3 | 25.5 | .97 |
| | 912.9 | 97.6 | .90 |
| Nutrition Standards | 501.3 | 155.0 | .76 |
| | 679.4 | 132.1 | .84 |
| Physical Education and Physical Activity | 662.1 | 96.4 | .85 |
| | 421.3 | 63.3 | .87 |
| Evaluation | 1147.0 | 122.7 | .90 |
| | 1511.2 | 87.4 | .93 |

* Proportion of the total variance occurring between LWPs.
† Convergence criteria for the model were met. The estimated variance component was 0.
One participant in the pilot consistently assigned higher scores than any of the 4 other contractors. This suggests that additional training on the instrument and ongoing monitoring of scores would be needed for projects that would require combining WellSAT scores from different raters. However, even with data from the overly positive rater, WellSAT demonstrated adequate psychometric properties to support its use as an assessment tool for program evaluation. To address the potential for rater effects during the use of WellSAT in Healthy Schools NY structured training was provided prior to field testing. This training had the added benefit of developing procedures to assist users in downloading and submitting WellSAT data files for aggregation.

During field testing several coordinators reported being asked to complete WellSAT as a group using printed copies of the tool and rating guidance. These requests provide additional support for WellSAT being accessible and well accepted by the school community. Ongoing analysis of data collected through the Healthy Schools NY will enable us to identify whether systematic differences in scores emerge when WellSAT is completed by a committee.

**IMPLICATIONS FOR SCHOOL HEALTH**

The pilot and field testing completed suggest WellSAT represents a feasible and reliable tool for health and school agencies to use in assessing the policies included in LWPs and improving their content to align with recognized standards.\(^1\)\(^,\)\(^7\)\(^,\)\(^8\) Our experience suggests that state and local agencies anticipating using the WellSAT with multiple users should consider providing structured training on WellSAT to assist new users and ensure consistent use. The WellSAT was well accepted by the 18 public health practitioners using the tool and the preliminary reaction from school districts and local wellness committees using WellSAT has been positive. In addition to using WellSAT for program evaluation, the NYSDOH and New York State Education Department (NYSED) are using WellSAT to rate all the school districts LWPs the NYSED has collected as part of ongoing administration and monitoring requirements associated with the Child Nutrition and WIC Reauthorization Act. The data from this larger coding project are intended to provide a context for interpreting the WellSAT ratings generated as part of Healthy Schools NY and establish a data source for ongoing public health surveillance and research opportunities to improve school health through strengthening LWPs.

**Human Subjects Approval Statement**

This study was determined to be exempt by the New York State Department of Health’s Institutional Review Board.

**REFERENCES**

the Gap Program, Health Policy Center, Institute for Health Research and Policy, University of Chicago; 2010.