



Published in final edited form as:

Health Place. 2020 November ; 66: 102388. doi:10.1016/j.healthplace.2020.102388.

A systematic review of existing observational tools to measure the food and physical activity environment in schools

Hannah G. Lane^{a,*}, Hannah G. Calvert^b, Rachel Deitch^c, Ryan Harris^d, Oyinlola T. Babatunde^e, Lindsey Turner^b, Erin R. Hager^c, Stephanie Jilcott Pitts^f

^aDuke University School of Medicine, Department of Population Health Sciences, Durham, 27701, NC, USA

^bBoise State University, College of Education, Boise, ID, USA

^cUniversity of Maryland School of Medicine, Department of Pediatrics, Division of Growth and Nutrition^a, Baltimore, MD, 21201, USA

^dUniversity of North Carolina Charlotte, Charlotte, NC, USA

^eEast Carolina University, Department of Nutrition Science, College of Allied Health Sciences, Greenville, NC, 27834, USA

^fEast Carolina University, Department of Public Health, Brody School of Medicine, Greenville, NC, 27834, USA

Abstract

Objective measures of schools' wellness-promoting environments are commonly used in obesity prevention studies; however, the extent to which findings from these measures translate to policy and practice is unknown. This systematic review describes the comprehensiveness and usability (practicality, quality, and applicability) of existing objective measures to assess schools' food and physical activity environments. A structured keyword search and standard protocol in electronic databases yielded 30 publications reporting on 23 measures that were deemed eligible to include in this review. Findings provide details to guide researchers and practitioners in selecting the best tool for use in policy and program evaluations.

1. Background

Childhood obesity and associated inequities remain a public health problem. Schools are a logical target for obesity prevention interventions, as children and adolescents spend half of their waking hours in school for more than 180 days/year (Khambalia et al., 2012; Story et al., 2006). School-based interventions to improve policies and practices that support student wellness have demonstrated potential to reduce childhood obesity inequities, given that nearly all children can benefit (Micha et al., 2018; Singh et al., 2017; Olstad et al., 2017; Morton et al., 2016). Interventions that address schools' food and physical

*Corresponding author: hannah.lane@duke.edu (H.G. Lane).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.healthplace.2020.102388>.

activity environments (i.e., physical structures/facilities that provide nutritious food options, create physical activity opportunities, reduce sedentary time) have particular promise as a mechanism for changing student behaviors (Ip et al., 2017; Sallis and Glanz, 2009; Story et al., 2009).

With this shifting focus to social and physical contexts of school, there has been increasing interest among researchers, practitioners, and policy advocates in evaluating food and physical activity environments. Such methods can examine whether policy, systems, and environmental change interventions to change school environments are being implemented as intended and are effective, and, importantly, can generate school-specific data to continue to inform improvements (Morton et al., 2016; Glanz et al., 2015). Methods to measure school contexts range from administrator surveys or semi-structured interviews to observational audits or checklists. Objective, observational methods such as audits are advantageous in that they measure the actual, rather than the perceived, school environment, thus reducing reporting bias.

Several systematic reviews have advanced our understanding of assessing nutrition and/or physical activity environments, but few have focused specifically on tools appropriate for use in a school setting (Saluja et al., 2018; Ajja et al., 2015; Mckinnon et al., 2009; Lytle and Sokol, 2017). Furthermore, while most reviews describe the psychometric properties of these measures (e.g., reliability, validity), descriptions of comprehensiveness and usability (e.g., practicality of administration, applicability across settings and populations, relevance to stakeholders) of measures is also important. To further advance our understanding of schools' food and/or physical activity environments, and subsequently improve selection and use of existing tools for research, school improvement, and/or policy decision-making, it is critical to identify and disseminate observational tools that are simple, practical, relevant and adaptable for use by various stakeholders (Glanz et al., 2015; Glasgow and Riley, 2013; Saluja et al., 2018).

Given the need for pragmatism and generalizability of these tools, this systematic review aims to describe the (1) comprehensiveness of currently available, research-tested observational tools that assess schools' environments for food and/or physical activity and (2) usability of those tools, specifically practicality, quality, and applicability.

2. Methods

This systematic review follows PRISMA guidelines (Moher et al., 2009). This review uses similar inclusion/exclusion criteria as a recent systematic review of audit tools for child-serving settings (Ajja et al., 2015), and builds upon that review by extracting information about tools' comprehensiveness and usability, focusing specifically on schools. Additionally, this review aims to identify research gaps and provide recommendations for the field.

2.1. Search strategy

The study authors, including a research librarian, developed and ran searches in PubMed, CINAHL (EBSCO), Scopus (Elsevier), SocIndex (EBSCO), and ERIC (Educational Resources Information Center) (EBSCO) (See Supplemental File 1). Specific search

strategies were created for each database and used a combination of both controlled vocabulary (organized words or phrases used to index and retrieve content) and keyword searching. The concept of “schools” included, but was not limited to: *schools, kindergarten, elementary school, middle school, high school*. Controlled vocabulary was not exploded because in many instances it generated search results related to higher education and college, which were not our focus. The concept for “audit” included: *audits, observation, checklist, inventory*. School environment concepts included but were not limited to: *nutrition, food, wellness, exercise, schoolyard, playground*. Truncation to search terms was used when appropriate, as well as title and abstract searching. Searches were originally completed in November 2016, then re-conducted in August 2018. Additional articles were identified from the reference lists of eligible articles and reviewed if appropriate based on the inclusion criteria.

2.2. Inclusion and exclusion criteria

Inclusion criteria were developed to capture tools that objectively measure the structural or physical environments of schools related to food and/or physical activity. No time limits were placed on the search. Articles were excluded if they: (1) were not published in peer-reviewed journals; (2) were not published in English; (3) referred to tools used outside of K-12 schools or exclusively in after school; or (4) did not include an observational measure of the structural or physical environment that specifically pertained to food and/or physical activity (e. g., behavioral observation tools). Articles were excluded if they did not include a thorough description of items and/or the tool could not be acquired through communication with authors. Based on a previously described strategy (Lewis et al., 2015), we contacted corresponding authors via email on three separate occasions to request tools and supplemental materials. The method by which the tool was accessed (e.g., contained in article, publicly available, or via author communication) is described in Table 1. Finally, articles were excluded if they described checklists (e.g., pre-defined list of visible items, like items in a vending machine) or inventories (e.g., listing types of competitive foods sold in school stores). The availability and potential for broad use of checklists/inventories is described in detail elsewhere (Lytle and Sokol, 2017).

2.3. Tool selection

After both searches (November 2016, August 2018) were conducted, initial titles were screened by four authors (HL, RD, EH, LT) to eliminate non-peer reviewed articles. Two research assistants then conducted the title and abstract review, with disputes resolved by a third reviewer (HL). A full text review was conducted by three authors (HL, HC, RD). Each article was reviewed for inclusion, then coded by two of the three authors, who met after coding to resolve discrepancies. If they could not reach consensus, the third author reviewed the article and provided a final judgement. For eligible articles, an email was sent to the corresponding author to request any additional publicly available, non-proprietary documents that may contain information on tool comprehensiveness and usability.

2.4. Data extraction

The data extraction protocol was developed iteratively by the authors. The protocol drew upon gaps identified in previous reviews of school audit tools, and existing frameworks

describing characteristics of pragmatic, useful measures (Saluja et al., 2018; Ajja et al., 2015; Lewis et al., 2015; Glasgow and Riley, 2013). The entire authorship team drafted the initial protocol, pilot-coded one article, then made refinements. Three authors (HL, HC, RD) coded an additional six articles and discussed them as a team before finalizing the extraction protocol.

The extraction protocol was derived from two existing frameworks: Glasgow and Riley's criteria for pragmatic measures and Lewis et al.'s evidence-based assessment. Glasgow and Riley (2013) cited the gap in translation from research to effective public health practice as partially attributed to the impracticality of measures. They suggested criteria to increase the applicability and pragmatism of measures (e.g., important to stakeholders, low burden for respondents and staff). Lewis et al. (2015) provided a detailed review of various measures related to a core set of seven implementation outcomes (e.g. acceptability, fidelity). These frameworks provide criteria for measures that are not just sufficiently rigorous to address a research question, but also practical enough to be useable in real-world settings and to provide recommendations for policy, intervention, and evaluation. For this review, the protocol was divided into two concepts, comprehensiveness and usability, and constructs were adapted based on the unique needs of environmental measures that may be used for monitoring implementation processes and outcomes. The protocol considered the context of existing best practices related to healthy eating and physical activity during school, and with the understanding that these were organizational, rather than individual-level measures (CDC, 2017; WHO, 2017 Alliance For A Healthier Generation, 2016; USDA, 2010; USDA 2016).

Comprehensiveness information for each tool included number and type of settings/domains (relevant areas of the school, which correspond to best practices) covered (see Table 2). Usability information consisted of fifteen items, which were extracted from each study. Six items described practicality, including the time, resources, and training/expertise required for administration. An additional four items described quality. While quality is typically defined by psychometric properties of reliability and validity (or trustworthiness, for qualitative data), our expanded definition includes describing the development and/or adaptation process and scoring protocol. Finally, five items described applicability, defined as the potential use for future research and school improvement, and policy/practice relevance. The applicability construct also included any "calls to action" or methods for disseminating results with stakeholders. Each construct was selected to provide tool users (practitioners and researchers) with information that would help determine which tool was suited for the users' available resources and purpose of the research or evaluation study, with a focus on translation from research/evaluation to practice. Each item was coded as yes (described) or no (not described). Items are defined in Tables 3–5. Details were extracted from the publication(s), with additional information acquired when the training manual or user's guide was accessed.

3. Results

3.1. Search results

A total of 4691 unique articles were initially identified through the electronic search. After title and abstract review, 243 articles were reviewed for eligibility. A total of 43 articles describing 30 studies were included (Fig. 1). These studies included use of 23 eligible tools.

3.2. Study and tool description

Table 1 lists the 30 studies, including tool name, purpose, setting and school type(s) where it was used, and types of items, as well as how actual tools and other documents such as user manuals were accessed. Studies most commonly occurred in elementary/primary schools ($n = 12$; 40%) or both elementary/primary and middle schools ($n = 3$; 10%), with 17 (56.7%) in the United States. Half ($n = 15$; 50%) of studies included schools serving predominantly racial/ethnic minority and/or socioeconomically disadvantaged student body.

Most ($n = 20$; 87%) of the 23 tools were used in a single study. Two tools, the Physical Activity Resource Assessment (PARA) and the COMPASS School Environment Application (CO-SEA), were used in two different studies. Another tool, the Sport, Physical Activity and Eating behavior: Environmental Determinants in Young people (SPEEDY) outdoor grounds audit, was used in six studies (see Table 1).

3.3. Comprehensiveness of 23 rated tools

Comprehensiveness is described in Table 2. The 23 tools covered an average of 2.7 ($SD = 2.0$) settings/domains. Twelve of these tools covered primarily food-related domains, five covered primarily physical activity-related domains (with water sources included), and six covered both. The cafeteria was the most common domain ($n = 12$), followed by outdoor play areas ($n = 11$). The least common were garden/landscape ($n = 6$) and indoor play areas ($n = 5$), and no tools assessed classrooms. Of the 23 tools, eight covered only one domain while one tool covered all but one (Webber et al., 2007). Most studies ($n = 21$, 70%) augmented the observational audit with other school-level data collection methods (e.g., administrator or food service interviews/surveys, Geographic Information Systems mapping).

3.4. Usability information reported in 30 studies

Information pertaining to usability (practicality, quality, and applicability) is described for all 30 studies reviewed (Tables 3–5). Two tools were adapted across multiple studies—CO-SEA, which was adapted from Canada to Guatemala, and SPEEDY, a tool developed in England and used in six studies. This adaptation process provided information on usability; thus, all 30 studies describing 23 tools are included.

Practicality.—The mean score for practicality was 3.6 ($SD = 1.4$), based on six criteria (Table 3). Most studies ($n = 26$; 87%) reported the number or range of items on the tool that assessed the structural environment. The data collection method (e.g., paper vs electronic) and length of time to complete the tool were less frequently reported (both $n = 12$; 40%). Some studies reported number of raters needed per school ($n = 23$; 77%), rater expertise

(n = 23; 77%) and training (n = 12; 40%). Among studies that reported length of time to complete the tools, times ranged from 10 min to 1–2 full days, and it was not always clear how much of that time was spent on the observational component if the tool included other data collection methods. The original CO-SEA was the only study to report using a software application rather than pencil/paper (Leatherdale et al., 2014a), and three studies reported utilizing photographs for data collection (Spruance et al., 2017; Black et al., 2015; Leatherdale et al., 2014a).

Of the 23 studies reporting on the number and expertise of raters who completed the tool at each school, most indicated 1–2 raters per school (often with the second to establish inter-rater reliability) and that raters were members of the research team. Three studies reported training non-research staff to complete the tool, including school personnel (John et al., 2016a,b), members of the community-academic partnership (Patel et al., 2009), and national fieldworkers (Laurie et al., 2017). The quality and duration of training varied. Common training strategies included certification and/or structured training using multiple methods (e.g., online and in-person), and supervised auditing.

When training manuals were available, these provided additional details regarding practicality. Manuals for the PARA (Lee et al., 2005), SPEEDY (Jones et al., 2010), ISAT (Broyles et al., 2015), and Environmental Assessment of Public Recreation Spaces (EAPRS) (Saelens et al., 2006) included a list of operational definitions to help with scale items. The ISAT manual also included a list of “quality control tips” that were likely refined throughout tool administration (Broyles et al., 2015). The Great Recess Framework Observational Tool (GRF-OT) included video examples to help auditors assess scaled items (Massey et al., 2018b). The School Food Environment Scan (Spruance et al., 2017), CO-SEA (Leatherdale et al., 2014a), and SF-EAT (Black et al., 2015) explicitly requested accompanying photographs. Manuals sometimes contained additional practical tips or instructions on navigating the school staff and environment, such as what to say to skeptical cafeteria or office staff (Spruance et al., 2017; Black et al., 2015), how to request a student or staff consultation to confirm observations (Katzmarzyk et al., 2013), or strategies for remaining objective (Ritchie et al., 2015).

Quality.—Mean score for reporting of quality was 2.3 (SD = 1.1) out of four criteria. Details for each study are included in Table 4. Most (n = 24; 80%) described some form of pilot or formative testing of new tools and/or a process of adapting existing tools for a new setting, population, or use. Information about reliability and validity of tools in their designated setting/population was reported in 14 (47%) and seven (23%) studies, respectively. Inter-rater reliability was the most common form of reliability and was reported as adequate in most studies. Types of validity reported included face validity, construct validity (by comparing audit scores to child physical activity levels), and content validity (by assessing fit using structural equation modeling). One study that triangulated data from the observational tool with field notes, the Food Decision Environment Tool study, reported establishing trustworthiness (Ozturk et al., 2016). Trustworthiness is a qualitative research concept to establish credibility, transferability, dependability, and confirmability of methods (Lincoln and Guba, 1985).

Most studies ($n = 23$; 77%) published at least some information about scoring, although the level of detail varied. Reported scoring methods ranged in complexity, including straight scoring (e.g., creating a sum score), calculating means, weighting by mean, or developing new variables. In several studies, observational audit data was combined with other data sources (e.g., principal surveys, field notes) to create categorical variables (Thomas et al., 2016a,b; Belansky et al., 2013; Rushing and Asperin, 2012). While most tools were scored using quantitative methods, a few tools relied on qualitative scoring methods, including thematic analysis and the constant comparative method (Ozturk et al., 2016; Black et al., 2015; Belansky et al., 2013; Rushing and Asperin, 2012; Patel et al., 2009).

The level of detail provided on the tool development/adaptation process varied across studies. Several described the systematic involvement of school stakeholders (e.g., cafeteria staff) during development to ensure relevancy and/or establish face validity (Massey et al., 2018a; Ozturk et al., 2016; Patel et al., 2009). Adaptations included: adapting neighborhood or park audits for a school setting (Findholt et al., 2011; Colabianchi et al., 2011), customizing items to be appropriate for a new population or policy environment (Tarun et al., 2017; Broyles et al., 2015; Briefel et al., 2009), or combining multiple existing tools (John et al., 2016a,b; Leatherdale et al., 2014a). Some studies provided additional quality-related details on the development and use of their scoring protocols. In the SF-EAT study, authors collaborated with community and school stakeholders to develop a tool and scoring protocol and analyze the data collected (Black et al., 2015). The SPAN-ET study reported the time commitment for scoring (2 h), and provided thorough scoring instructions (John et al., 2016a,b). The ISAT study explicitly described that its rationale for developing the scoring system a priori was to increase objectivity of coders (Broyles et al., 2015), whereas the School Lunchroom Audits tool scoring system was intentionally developed after data collection, to allow the data to define the constructs so that none were missed (Thomas et al., 2016a,b). Belansky et al. (2013) described a systematic method to ensure that findings would be actionable, using literature evidence to classify the school-level changes reported in their study as “effective,” “promising” or “emerging” based on likelihood of improving behavior and weight outcomes.

Applicability.—Mean score for applicability was 2.7 ($SD = 1.4$) out of five criteria (Table 5). Most studies ($n = 23$; 83%) described how tools linked to implementation of a specific policy, program, or intervention. Limitations related to the tools themselves or their use were described across most ($n = 25$; 83%) studies. Less than half of studies reported potential future tool uses for researchers, policy makers, or school stakeholders ($n = 12$; 40%) and/or provided a “call to action” for schools based on the data ($n = 18$; 55%). Only five (17%) reported sharing findings with school stakeholders during or after the research study.

The purpose and scope for the use of each tool varied broadly, ranging from assessing implementation of federal policies to measuring impact of small-scale interventions. When studies described limitations, they mostly pertained to use of the tool, rather than the tools themselves. Study limitations included lack of validity testing or small sample sizes for validity testing, cross-sectional administration, and limited generalizability. The version of CO-SEA (developed in Canada) used in Guatemala cited a lack of tailoring as a limitation, because data related to environmental and cultural differences were not captured (Godin et

al., 2017). A longitudinal study that used SPEEDY to assess both elementary and middle schools pointed out that the instrument, which was tested for reliability in elementary schools but not tailored for middle schools, may not capture relevant differences between the two (Harrison et al., 2016a,b). Authors commonly described limitations related to using the tool in single point-in-time observation studies, as they may not reflect the dynamic aspects of policy or program implementation (Thomas et al., 2016a,b; Black et al., 2015; Ritchie et al., 2015; Van Der Horst et al., 2008; Dias et al., 2017; Colabianchi et al., 2011). Some studies discussed the limitations of the tools themselves rather than their use in the studies, including rating scales used to assess “condition” of facilities or equipment, which authors noted can be somewhat subjective and thus vulnerable to measurement error (Leatherdale et al., 2014a; Jones et al., 2010; Colabianchi et al., 2011).

Potential future uses of tools were primarily research-focused, and included expanding to other school types and regions to improve generalizability (Massey et al., 2018b; John et al., 2016a,b; Godin et al., 2017; Leatherdale et al., 2014a; Colabianchi et al., 2011), identifying future intervention targets (Ozturk et al., 2016), or developing process evaluation metrics (Newton et al., 2011). A few studies recognized potential uses by school stakeholders to monitor school-level policy implementation (John et al., 2016a,b; Patel et al., 2009) or to conduct a needs assessment (Black et al., 2015). When studies provided calls to action for schools and/or districts, they included the need to increase awareness, training, and buy-in from staff and students, and/or provided recommendations for structural, policy, or marketing changes related to the domains covered. The five studies that reported disseminating results noted that potential audiences include community advisory boards, school stakeholders, administrators, and school-based extension educators (Chalkley et al., 2018; Godin et al., 2017; John et al., 2016a,b; Leatherdale et al., 2014b; Patel et al., 2009). Only one study explicitly described the platform for stakeholder dissemination (e.g., written report, in-person conversation) and provided templates for reports in the training manual (John et al., 2016a,b). Data from the CO-SEA tool (along with other data sources) were used to generate tailored profiles, which were given to schools along with contact information for technical assistance providers (Leatherdale et al., 2014b; Godin et al., 2017). Patel et al. (2009) indicated that findings were shared with the community-academic partnership to improve implementation of a local cafeteria improvement policy, and described the specific initiatives undertaken as a result.

4. Discussion

This systematic review describes 30 studies that used 23 unique, peer-reviewed observational tools for scientific research to assess the food and/or physical activity environment of schools. While some of these tools have been identified in previous reviews (Saluja et al., 2018; Ajja et al., 2015), ours uniquely identifies tools’ wide-ranging uses and characteristics, and reports on their comprehensiveness and usability. Taken together, the studies in this review demonstrate the broad utility of school observational audit tools for many purposes, with information to assist researchers and practitioners in selecting a tool most appropriate for their purpose. Results also provide priorities for future tool developers and/or users regarding tools’ comprehensiveness, practicality, quality and applicability.

4.1. Recommendations for future tool developers and/or users

In terms of comprehensiveness, tools ranged in their coverage of domains associated with policies and best practices related to food and/or physical activity. Two specific domains assessed by the majority of tools were the cafeteria and outdoor play areas. There are a few domains that could benefit from additional measures development, including indoor play areas and classrooms. Characteristics of classrooms—such as the availability of adjustable-height desks (i.e., those that encourage standing instead of sitting) or exercise balls/chairs—are associated with less sedentary behavior among students (Hinckson et al., 2016), but no standardized tools were identified to measure these characteristics. Reliable and valid measures would allow researchers to evaluate interventions to conduct needs assessments and surveillance in these important, understudied domains.

Although the objective of this review was not to rank quality, we found that many tools have undergone rigorous psychometric testing. Reliability, validity and trustworthiness (for qualitatively-oriented tools) are critical, and they differentiate research-tested audit tools versus those that are primarily used for continuous improvement efforts. It is encouraging that inter-rater reliability was commonly reported; however, only one study reported assessing test-retest reliability (Massey et al., 2018b), which is also recommended (Gwet, 2014). Zenk et al. (2007) emphasize that assessing both forms of reliability can ensure that the operational definition is adequate, the condition is stable/clearly visible, and the raters are adequately trained. Future tool development should consider both test-retest and inter-rater reliability.

For categories of tools where there is no true existing gold standard, which is the case with environmental observations, validity testing presents a challenge. Researchers often explore weaker forms of validity, such as face validity (items, “on their face”, are logical translations of constructs) (Royal, 2016; Trochim, 2006). This was the most common type of validity reported in these studies. Two studies that used versions of the SPEEDY tool reported on construct validity (or convergent validity, the degree to which a measure relates to or converges on a construct it theoretically should relate to) by examining associations of the tool scores with high or low student physical activity behaviors (Trochim, 2006). Given that these tools are developed and used to assess hypotheses around environmental influences on behavior (i.e. understanding associations between the school environment and student behaviors), caution should be taken to differentiate between tool validation as compared to testing hypotheses about causal relationships between variables (environment and behavior).

Many tools employed qualitative methods, but only one reported on the quality of those methods using Lincoln and Guba’s trustworthiness concept (credibility, transferability, confirmability and dependability). While defining and operationalizing quality is debated among qualitative methodologists (Cypress, 2017; Leung, 2015), trustworthiness is a commonly cited method and, thus, serves as a useful starting point. Regardless of approach, qualitative tool developers should clearly articulate and provide justifications for their chosen quality standards across all stages of development, testing, scoring, and reporting (Lincoln and Guba, 1985; Qualitative Research In Implementation Science (Qualris) Group, 2019).

While establishing reliability, validity, and trustworthiness is essential for rigor and reproducibility, it is important to be mindful that these psychometric dimensions are often less important to school stakeholders than tools' cultural appropriateness or ability to generate understandable and actionable findings. In their criteria for pragmatic measures, Glasgow and Riley (2013) emphasize the importance of psychometrically strong measures; however, they do note that in practical application, "perfect is the enemy of good." This is particularly true of more-objective tools such as audits, which inherently have more stability. Thus, researchers selecting a measurement tool should consider giving equal weight to relevance as is typically given to rigor. Relevance can be achieved through a systematic adaptation process—or a systematic development process, in the case of tools to measure new domains. As such, the field would benefit from consensus guidelines about ensuring consistency, rigor, and relevance, and publishing details about tool development and adaptation processes.

For some purposes, such as studies with a wide geographic scope, use of lower-burden strategies may be perceived as preferable to on-site visits. However, efforts to ensure practicality can increase the utility of observational tools which may require more time and staffing to gather data. Transparency and clarity about resources, expertise, and time required to carry out observational audits can improve uptake by researchers as well as educators and community partners. Some approaches may minimize data collection time and resources needed, such as creating a phone application or electronic tool for data capture, and collecting photos as supporting evidence—similar to the CO-SEA tool (Leatherdale et al., 2014a). This may reduce data entry errors and enable swifter scoring, interpretation, and sharing. However, the trade-off could be challenges related to upkeep and required upgrades. Another option would be to provide "brief" versions that are less burdensome. In the SF-EAT study, researchers and school stakeholders created a brief version of the thorough, mixed methods tool that assessed needs at baseline in order to monitor changes over time (Black et al., 2015). As recommended by Katzmarzyk et al. (2013), efficiency, comprehensiveness, and relevance can be improved by including stakeholders in data collection. One consideration when adapting or developing measures is the trade-off between comprehensiveness and practicality. Similar to Saluja et al. (2018), we found a lack of comprehensive instruments: only one was identified, and it was developed with teachers in mind (e.g., while it assessed indoor play areas, it only did so with consideration to how it might be used by teachers) (Webber et al., 2007).

In addition to describing the adaptation/development process and reliability and validity testing, another critical component of quality in our review was the description of scoring methods employed by various studies. Without understanding how a tool is scored, data may never be interpreted, nor be useful for stakeholders (Glasgow and Riley, 2013). We found that while most studies provided actual scores, very few provided protocols or instructions. Incorporating details on scoring into training guides, such as the instructions provided in SPAN-ET's user's guide, could improve quality by ensuring that data are collected reliably and correctly, and consistently scored and interpreted (John et al., 2016a,b). Even if a tool's scoring protocol is easy to use and publicized, it may still lack usability if it does not align with school or district priorities, and/or lacks actionable implementation steps (Glanz et al., 2015). The SF-EAT scoring protocol recognized the complexities of the concepts being

assessed and thus, drew upon a systems theory approach, with scores compiled from various stakeholders on a given construct. This process was strengthened by including individuals with first-hand knowledge of each school in the scoring process (Black et al., 2015). Patel et al. (2009) described a community-based participatory research approach to developing the tool and scoring system. The tool described by Belansky et al. (2013) was scored within the context of Brennan's Evidence Typology, which classified the environmental or policy changes detected by the tool as "effective," "promising," or "emerging" based on their potential to have clinically meaningful impact. Describing not just the change that occurred but also its potential to impact student behavior can enable schools to better plan for resource allocation. The School Nutrition Environment Observation (SNEO) used a simple protocol that involved sorting items into "recommended" (e.g., offers fruits and vegetables) versus "non-recommended" (e.g., presence of concessions) based on evidence (Newton et al., 2011). These tools demonstrate the potential for researchers and school stakeholders to work together to create high quality, well-described, actionable scoring systems that meet stakeholders' needs.

An important future direction is to consider existing tools' ability to detect changes over time as well as variability across schools. Sensitivity to change over setting and time is a hallmark of a pragmatic, useable measure (Glasgow and Riley, 2013), and is particularly important when recommendations and policies frequently change or evolve, as has been the reality for school nutrition and physical activity policy and practice. The SPAN-ET developers recognized the importance of developing an adaptable tool, and thus provided a simple, item-by-item scoring system to allow for revisions or additions if best practices change (John et al., 2016a,b). Tools should also be designed and tested to be able to detect whether evidence-based practices are being implemented or whether interventions focused on improving environments work. Additionally, to better tailor future policies and implementation of evidence-based practice, tools should be able to detect differences in implementation between schools, particularly between school types (e.g., elementary/primary, middle, high).

Similar to the rest of the literature in this area, the studies in our review primarily focused on contributions to research, rather than practical implications for schools. In the few cases when study authors described recommendations for schools based on their findings, it was not often clear that the schools/districts involved in the research were informed of these recommendations. We can improve applicability of tools for policy and practice in several ways. First, in addition to refining the information reported in scientific manuscripts (or supplemental materials), authors should ensure that tools and supporting materials are publicly available and that they thoroughly describe administration, scoring, and dissemination. Such materials could be provided and/or described on study protocol websites, available via email, or submitted to a federally funded website or national registry. (Glanz et al., 2015). For example, the Smarter Lunchrooms tool, PARA, and SPAN-ET are all available in the 2016 SNAP-Ed toolkit (<https://snapedtoolkit.org/>). We found that only a few studies reported providing results to schools or districts. School stakeholders should be involved in designing the dissemination process from the beginning of a study, not only by providing data but also by determining who should receive results and in what format. Most studies reported stakeholder involvement to some extent, but very few

reported their involvement throughout development/adaptation, administration, scoring, and dissemination. According to Glasgow and Riley (2013), continuous stakeholder involvement is another hallmark of a useable, practical measure, as it increases the likelihood that findings will be applicable for policy and practice, and improves the practicality and quality of the data collection process. We recognize that the indicators related to applicability do not necessarily align with what is traditionally reported to journals or considered important components of scientific manuscripts. Indeed, it is possible that authors were more intentional about items related to “usability” but did not prioritize their discussion in publications. As such, we echo calls of other obesity prevention researchers to prioritize reporting of indicators like applicability, often referred to as “external validity” (Glanz et al., 2015; Klesges et al., 2008; Glasgow, 2008). With this review, we hope to generate reporting norms for measurement studies to include school stakeholder involvement and dissemination of findings. Reporting on the 15 usability indicators described in this review could encourage prioritization of external validity, and thus improve the translation of findings regarding school environments into actionable policies and practices.

4.2. Focusing on generalizability and understanding inequities

The studies reviewed occurred across a variety of settings, populations, and school types (e.g., elementary/primary, middle and high), with several studies using the same tools in different countries or school types. About half of the studies focused on schools serving socially disadvantaged students. The focus on the school environment across all school types is important, given that changes in school environments between elementary/primary and middle school are predictive of less-healthy behaviors as students get older (Harrison et al., 2016; Black et al., 2015; Marks et al., 2015). While a common global measure may not be possible due to differences in schools/districts and across countries, we identified several studies that showed that adaptation can happen successfully. If future studies adapt existing tools rather than developing new ones, the field could address lack of generalizability and small sample size (limitations noted in most reviewed studies), as well as strengthen tools’ ability to detect changes across locations. The six studies that used the SPEEDY tool across different countries and school types demonstrate the broad utility of the tool. These authors could conduct a pooled analysis with a variety of school types, locations, and policy environments to improve our understanding of generalizability and sensitivity to between-school differences.

Finally, we note that there was limited information about how tools can identify and address inequities in school environments. Although some studies occurred in socially disadvantaged communities, few described whether the tools are appropriate across a variety of school demographics. The availability of practical tools is particularly important for schools in these disadvantaged communities, because they are often under-resourced and over-burdened. Describing these inequities will ideally result in the intentional allocation of additional resources to improve school environments and thus, reduce inequities (Taber et al., 2015; Carlson et al., 2014). Moreover, the implementation of tools to develop an evidence-based understanding of school environment components either associated with or that impact student health behaviors among at-risk children could inform school building design decisions, particularly in lower-resource communities (Gorman et al., 2007).

4.3. Limitations of this review

There are several important limitations of this review. While we made several attempts to contact authors, we excluded 17 studies due to no response; thus, the review is not exhaustive, but it reflects what may be available for other users. An additional limitation was that, similar to other systematic reviews, relevant articles may have been published between the conclusion of our search and publication. However, our search yielded 30 high-quality articles which allowed us to make strong conclusions regarding 23 useable, practical and applicable tools to improve the school food and physical activity environments. Our review includes only what was reported in articles and other materials provided or publicly available. Given that articles varied in their purpose (e.g., tool validation versus large intervention trial) and differed in the space devoted to describing the tool, it is possible that we missed information, such as interpretability of scoring or whether and how findings were shared with schools, which was either known but not reported or not yet known. To overcome this limitation, we contacted all authors to request additional information that was publicly accessible but not available in the peer-reviewed literature. Finally, while there are many methods of assessing the school environment, these findings only pertain to observational audit tools or the observational component of tools. Non-observational methods (e.g., key informant interviews, administrator surveys, policy documents) provide different but important information, which is sometimes overlapping. Thus, it is possible that, when tools utilized observational and non-observational methods to assess the same domains, our extraction protocol may have underestimated the items covered by the observational component.

4.4. Conclusions

This review provides a list of tools for assessing school environments relevant to food and/or physical activity, the areas of the school they cover, and extensive information about the usability of each tool. We considered each tool's usability, but found that such information was not reported often. We urge researchers using these tools or developing new tools to report usability characteristics. This review can assist researchers and practitioners in selecting appropriate tools for research, intervention planning, and evaluation of efforts to improve schools' food and physical activity environments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was support by the Centers for Disease Control's NOPREN (Nutrition and Obesity Policy Research and Evaluation Network) and PAPRN (Physical Activity Policy Research Network) Networks; National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases [F32DK115146] and National Heart, Lung, and Blood Institute [K12HL13830]; United States Department of Education, Institute of Education Sciences [R305A150277]. United States Department of Agriculture (USDA) Agriculture and Food Research Initiative Grant [30000045856] [2016-68001-24927].

References

- Adachi-Mejia AM, L MR, Skatrud-Mickelson M, Li Z, Purvis LA, Titus LJ, Beach ML, Dalton MA, 2013. Variation in access to sugar-sweetened beverages in vending machines across rural, town and urban high schools. *Publ. Health* 127, 485–491.
- Ajja R, Michael WB, Chandler Jessica, Kaczynski Andrew T., Ward Dianne S., 2015. Physical activity and healthy eating environmental audit tools in youth care settings: a systematic review. *Prev. Med* 77, 80–98. [PubMed: 25964078]
- Alliance For A Healthier Generation, 2016. Healthy Schools Program framework of best practices Accessed September 2016. https://www.healthiergeneration.org/_asset/1062yk/07-278_Hspframework.pdf.
- Belansky ES, Nick C, Chavez Robert, Crane Lori A., Waters Emily, Marshall Julie A., 2013. Adapted intervention mapping: a strategic planning process for increasing physical activity and healthy eating opportunities in schools via environment and policy change. *J. Sch. Health* 83, 194–205. [PubMed: 23343320]
- Black JL, V CE, Ahmadi N, Chapman GE, Carten S, Edward J, Shulhan S, Stephens T, Rojas A, 2015. Sustainability and public health nutrition at school: assessing the integration of healthy and environmentally sustainable food initiatives in Vancouver schools. *Publ. Health Nutr* 18, 2379–2391.
- Briefel RR, C MK, Cabili C, Wilson A, Gleason PM, 2009. School food environments and practices affect dietary behaviors of Us public school children. *J. Am. Diet Assoc* 109, S91–S107. [PubMed: 19166677]
- Broyles ST, Drazba KT, Church TS, Chaput JP, Fogelholm M, Hu G, Kuriyan R, Kurpad A, Lambert EV, Maher C, Maia J, Matsudo V, Olds T, Onywera V, Sarmiento OL, Standage M, Tremblay MS, Tudor-Locke C, Zhao P, Katzmarzyk PT, 2015. Development and reliability of an audit tool to assess the school physical activity environment across 12 countries. *Int. J. Obes. Suppl* 5, S36–S42. [PubMed: 27152183]
- Carlson JA, Mignano AM, Norman GJ, Mckenzie TL, Kerr J, Arredondo EM, Madanat H, Cain KL, Elder JP, Saelens BE, Sallis JF, 2014. Socioeconomic disparities in elementary school practices and children’s physical activity during school. *Am. J. Health Promot* 28, S47–S53. [PubMed: 24380465]
- Centers For Disease Control And Prevention and Shape America—Society of Health and Physical Educators, 2017. Centers for Disease Control and Prevention and Shape America—Society of Health and Physical Educators. *Strategies for Recess in Schools* Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2017, Atlanta, Ga.
- Chalkley AE, Routen AC, Harris JP, Cale LA, Gorely T, Sherar LB, 2018. Marathon Kids UK: study design and protocol for a mixed methods evaluation of a school-based running programme. *Bmj Open* 8, e022176.
- Colabianchi N, M AL, Swayampakala K, 2011. Features and amenities of school playgrounds: a direct observation study of utilization and physical activity levels outside of school time. *Int. J. Behav. Nutr. Phys. Activ* 8, 32-5868-8-32.
- Cypress BS, 2017. Rigor or reliability and validity in qualitative research: perspectives, strategies, reconceptualization, and recommendations. *Dimens. Crit. Care Nurs* 36 (4), 253–263. [PubMed: 28570380]
- Dias AF, Lemes VB, Brand C, Mello JB, Gaya AR, Gaya ACA, 2017. Association between school structure and physical activity in physical education class and school recess. *Revista Brasileira de Cineantropometria e Desempenho Humano* 19, 164–173.
- Findholt NE, M YL, Jerofke LJ, Brogoitti VW, 2011. Environmental influences on children’s physical activity and eating habits in a rural Oregon County. *Am. J. Health Promot. : Ajhp* 26, e74–85. [PubMed: 22040399]
- Finkelstein DM, H EL, Whitaker RC, 2008. School food environments and policies in Us public schools. *Pediatrics* 122, e251–e259. [PubMed: 18595970]
- Fox MK, D AH, Wilson A, Gleason PM, 2009a. Association between school food environment and practices and body mass index of Us public school children. *J. Am. Diet Assoc* 109, S108–S117. [PubMed: 19166665]

- Fox MK, G A, Nogales R, Wilson A, 2009b. Availability and consumption of competitive foods in Us public schools. *J. Am. Diet Assoc* 109, S57–S66. [PubMed: 19166673]
- Glanz K, Sallis JF, Saelens BE, 2015. Advances in physical activity and nutrition environment assessment tools and applications: recommendations. *Am. J. Prev. Med* 48, 615–619. [PubMed: 25891061]
- Glasgow RE, 2008. What types of evidence are most needed to advance behavioral medicine? *Ann. Behav. Med* 35, 19–25. [PubMed: 18347901]
- Glasgow RE, Riley WT, 2013. Pragmatic measures: what they are and why we need them. *Am. J. Prev. Med* 45, 237–243. [PubMed: 23867032]
- Godin KM, Chacón V, Barnoya J, Leatherdale ST, 2017. The school environment and sugar-sweetened beverage consumption among Guatemalan adolescents. *Publ. Health Nutr* 20, 2980–2987.
- Gorman N, Lackney JA, Rollings K, Huang TT, 2007. Designer schools: the role of school space and architecture in obesity prevention. *Obesity* 15, 2521–2530. [PubMed: 18070739]
- Gwet KL, 2014. *Handbook of Inter-rater Reliability: the Definitive Guide to Measuring the Extent of Agreement Among Raters* Advanced Analytics, Llc, Gaithersburg, Md.
- Harrison F, V S, E.M., Corder K, Jones A, 2016a. School grounds and physical activity: associations at secondary schools, and over the transition from primary to secondary schools. *Health Place* 39, 34–42. [PubMed: 26922516]
- Harrison F, Van Sluijs EM, Corder K, Jones A, 2016b. School grounds and physical activity: associations at secondary schools, and over the transition from primary to secondary schools. *Health Place* 39, 34–42. [PubMed: 26922516]
- Hecht AA, Grumbach JM, Hampton KE, Hecht K, Braff-Guajardo E, Brindis CD, Mcculloch CE, Patel AI, 2017. Validation of a survey to examine drinking-water access, practices and policies in schools. *Publ. Health Nutr* 20, 3068–3074.
- Hinckson E, Salmon J, Benden M, Clemes SA, Sudholz B, Barber SE, Aminian S, Ridgers ND, 2016. Standing classrooms: research and lessons learned from around the world. *Sports Med* 46, 977–987. [PubMed: 26626071]
- Hunter S, L ST, Storey K, Carson V, 2016. A quasi-experimental examination of how school-based physical activity changes impact secondary school student moderate- to vigorous- intensity physical activity over time in the COMPASS study. *Int. J. Behav. Nutr. Phys. Activ* 13, 86-8016-0411-9.
- Hyndman B, Chancellor B, 2017. Are secondary school environments conducive for active play opportunities? An objective assessment across Australian secondary school playgrounds. *International Journal of Play* 6, 40–52.
- Ip P, Ho FK, Louie LH, Chung TW, Cheung YF, Lee SL, Hui SS, Ho WK, Ho DS, Wong WH, Jiang F, 2017. Childhood obesity and physical activity-friendly school environments. *J. Pediatr* 191, 110–116. [PubMed: 28987751]
- John DH, Gunter K, Jackson JA, Manore M, 2016a. Developing the school physical activity and nutrition environment tool to measure Qualities of the obesogenic context. *J. Sch. Health* 86, 39–47. [PubMed: 26645419]
- John DH, G K, Jackson JA, Manore M, 2016b. Developing the school physical activity and nutrition environment tool to measure Qualities of the obesogenic context. *J. Sch. Health* 86, 39–47. [PubMed: 26645419]
- Jones NR, J A, Van Sluijs EM, Panter J, Harrison F, Griffin SJ, 2010. School environments and physical activity: the development and testing of an audit tool. *Health Place* 16, 776–783. [PubMed: 20435506]
- Katzmarzyk PT, B TV, Broyles ST, Champagne CM, Chaput JP, Fogelholm M, Hu G, Johnson WD, Kuriyan R, Kurpad A, Lambert EV, Maher C, Maia J, Matsudo V, Olds T, Onywera V, Sarmiento OL, Standage M, Tremblay MS, Tudor-Locke C, Zhao P, Church TS, 2013. The international study of childhood obesity, Lifestyle and the environment (iscole): design and methods. *BMC Publ. Health* 13, 900-2458-13-900.
- Khambalia AZ, Dickinson S, Hardy LL, Gill T, Baur LA, 2012. A synthesis of existing systematic reviews and meta-analyses of school-based behavioural interventions for controlling and preventing obesity. *Obes. Rev* 13, 214–233. [PubMed: 22070186]

- Klesges LM, Dziewaltowski DA, Glasgow RE, 2008. Review of external validity reporting in childhood obesity prevention research. *Am. J. Prev. Med* 34, 216–223. [PubMed: 18312810]
- Laurie SM, Faber M, Maduna MM, 2017. Assessment of food gardens as nutrition tool in primary schools in South Africa. *S. Afr. J. Clin. Nutr* 30, 80–86.
- Leatherdale ST, Bredin C, Blashill J, 2014a. A software application for use in handheld devices to collect school built environment data. *Measurement* 50, 331–338.
- Leatherdale ST, Brown KS, Carson V, Childs RA, Dubin JA, Elliott SJ, Faulkner G, Hammond D, Manske S, Sabiston CM, Laxer RE, Bredin C, Thompson-Haile A, 2014b. The Compass study: a longitudinal hierarchical research platform for evaluating natural experiments related to changes in school-level programs, policies and built environment resources. *BMC Publ. Health* 14, 331.
- Lee RE, Booth KM, Reese-Smith JY, Regan G, Howard HH, 2005. The Physical Activity Resource Assessment (Para) instrument: evaluating features, amenities and incivilities of physical activity resources in urban neighborhoods. *Int. J. Behav. Nutr. Phys. Activ* 2, 13–13.
- Leung L, 2015. Validity, reliability, and generalizability in qualitative research. *J. Fam. Med. Prim. Care* 4 (3), 324–327.
- Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG, 2015. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implement. Sci* 10, 155. [PubMed: 26537706]
- Lewis L, M C, Katzmarzyk P, Olds T, 2016. Individual and school-level socioeconomic gradients in physical activity in Australian schoolchildren. *J. Sch. Health* 86, 105–112. [PubMed: 26762821]
- Lincoln YS, Guba EG, 1985. *Naturalistic Inquiry* Sage Publications, Inc, Beverly Hills, CA.
- Lytle LA, Sokol RL, 2017. Measures of the food environment: a systematic review of the field, 2007–2015. *Health Place* 44, 18–34. [PubMed: 28135633]
- Mantjes JA, J AP, Corder K, Jones NR, Harrison F, Griffin SJ, Van Sluijs EM, 2012. School related factors and 1yr change in physical activity amongst 9–11 year Old English schoolchildren. *Int. J. Behav. Nutr. Phys. Activ* 9, 153-5868-9-153.
- Marks J, Barnett LM, Strugnell C, Allender S, 2015. Changing from primary to secondary school highlights opportunities for school environment interventions aiming to increase physical activity and reduce sedentary behaviour: a longitudinal cohort study. *Int. J. Behav. Nutr. Phys. Activ* 12, 59.
- Massey WV, Stellino MB, Fraser M, 2018a. Individual and environmental correlates of school-based recess engagement. *Prev Med Rep* 11, 247–253. [PubMed: 30094127]
- Massey WV, Stellino MB, Mullen SP, Claassen J, Wilkison M, 2018b. Development of the great recess framework - observational tool to measure contextual and behavioral components of elementary school recess. *BMC Publ. Health* 18, 1–11.
- Mckinnon RA, R J, Morrisette MA, Lytle LA, Yaroch AL, 2009. Measures of the food environment. A compilation of the literature, 1990–2007. *Am. J. Prev. Med* 36, S124–S133. [PubMed: 19285203]
- Micha R, Karageorgou D, Bakogianni I, Trichia E, Whitsel LP, Story M, Penalvo JL, Mozaffarian D, 2018. Effectiveness of school food environment policies on children’s dietary behaviors: a systematic review and meta-analysis. *PloS One* 13, e0194555. [PubMed: 29596440]
- Moher D, Liberati A, Tetzlaff J, Altman DG, 2009. Preferred reporting items for systematic reviews and meta-analyses: the Prisma statement. *PLoS Med* 6, e1000097. [PubMed: 19621072]
- Morton KL, Atkin AJ, Corder K, Suhrcke M, Van Sluijs EMF, 2016. The school environment and adolescent physical activity and sedentary behaviour: a mixed-studies systematic review. *Obes. Rev. : an official journal of the International Association for the Study of Obesity* 17, 142–158.
- Newton RL, T JL, Rau KK, Ragusa SA, Sample AD, Singleton NN, Anton SD, Webber LS, Williamson DA, 2011. Psychometric characteristics of process evaluation measures for a rural school-based childhood obesity prevention study: Louisiana Health. *Am. J. Health Promot. : Ajhp* 25, 417–421. [PubMed: 21721969]
- Olstad DL, Ancilotto R, Teychenne M, Minaker LM, Taber DR, Raine KD, Nykiforuk CIJ, Ball K, 2017. Can targeted policies reduce obesity and improve obesity-related behaviours in socioeconomically disadvantaged populations? A systematic review. *Obes. Rev* 18, 791–807. [PubMed: 28434185]

- Ozturk OD, M MM, Blake CE, Frongillo EA, Jones SJ, 2016. Development of a structured observational method for the systematic assessment of school food-choice architecture. *Ecol. Food Nutr* 55, 119–140. [PubMed: 26654767]
- Pate RR, M KL, Colabianchi N, Troiano RP, Reis JP, Carroll DD, Fulton JE, 2015. Physical activity measures in the healthy communities study. *Am. J. Prev. Med* 49, 653–659. [PubMed: 26384937]
- Patel AI, B LM, Uyeda KE, Martinez H, Knizewski R, Ryan GW, Schuster MA, 2009. School site visits for community-based participatory research on healthy eating. *Am. J. Prev. Med* 37, S300–S306. [PubMed: 19896033]
- Qualitative Research in Implementation Science, 2019. Division of Cancer Control and Population Sciences, National Cancer Institute. Accessed May 6, 2020 from: <https://cancercontrol.cancer.gov/IS/docs/NCI-DCCPS-ImplementationScience-WhitePaper.pdf>.
- Ritchie LD, W P, Woodward-Lopez G, Thompson FE, Loria CM, Wilson DK, Kao J, Crawford PB, Webb KL, 2015. The healthy communities study nutrition assessments: child Diet and the school nutrition environment. *Am. J. Prev. Med* 49, 647–652. [PubMed: 26384936]
- Royal K, 2016. Face validity” is not a legitimate type of validity evidence! *Am. J. Surg* 1026–1027.
- Rushing KA, Asperin AE, 2012. Building the school nutrition program Brand personality within the school setting. *J. Child Nutr. Manag* 36.
- Saelens BE, Frank LD, Auffrey C, Whitaker RC, Burdette HL, Colabianchi N, 2006. Measuring physical environments of parks and playgrounds: eaprs instrument development and inter-rater reliability. *J. Phys. Activ. Health* 3, S190–s207.
- Sallis JF, Glanz K, 2009. Physical activity and food environments: solutions to the obesity epidemic. *Milbank Q* 87, 123–154. [PubMed: 19298418]
- Saluja K, Rawal T, Bassi S, Bhaumik S, Singh A, Park MH, Kinra S, Arora M, 2018. School environment assessment tools to address behavioural risk factors of non-communicable diseases: a scoping review. *Preventive Medicine Reports* 10, 1–8. [PubMed: 29868351]
- Singh A, Bassi S, Nazar GP, Saluja K, Park M, Kinra S, Arora M, 2017. Impact of school policies on non-communicable disease risk factors - a systematic review. *BMC Publ. Health* 17, 292-292.
- Spruance LA, Myers L, O'malley K, Rose D, Johnson CC, 2017. Individual- and school-level factors related to school-based Salad bar use among children and adolescents. *Health Educ. Behav* 44, 885–897. [PubMed: 28161990]
- Story M, Kaphingst KM, French S, 2006. The role of schools in obesity prevention. *Future Child* 16, 109–142. [PubMed: 16532661]
- Story M, Nannery MS, Schwartz MB, 2009. Schools and obesity prevention: creating school environments and policies to promote healthy eating and physical activity. *Milbank Q* 87, 71–100. [PubMed: 19298416]
- Taber DR, chriqui Jf, powell Lm, perna Fm, robinson Wr, chaloupka Fj, 2015. Socioeconomic differences in the association between competitive food laws and the school food environment. *J. Sch. Health* 85 (9), 578–586. [PubMed: 26201754]
- Tarun S, Arora M, Rawal T, Benjamin Neelon SE, 2017. An evaluation of outdoor school environments to promote physical activity in Delhi, India. *BMC Publ. Health* 17, 1–9.
- Thomas LN, H TF, Gaines A, Dollahite JS, 2016a. Implementing Smarter Lunchrooms Makeovers in New York state middle schools: an initial process evaluation. *Arch. Publ. Health* 74, 41.
- Thomas LN, H TF, Gaines A, Dollahite JS, 2016b. Implementing Smarter Lunchrooms Makeovers in New York state middle schools: an initial process evaluation. *Archives of public health Archives belges de sante publique* 74, 41. [PubMed: 27708776]
- Trochim WM, 2006. The Research Methods Knowledge Base, second ed Internet WWW page. URL: <http://www.socialresearchmethods.net/kb/>. (version current as of October 20, 2006).
- United States Department of Agriculture, 2010. Healthy hunger-free kids Act of 2010 Pub L No 111–296, 124 Stat 3183. 3183. <https://www.gpo.gov/fdsys/pkg/Plaw-111publ296/pdf/Plaw-111publ296.pdf>.
- United States Department of Agriculture, 2016. Local school wellness policy implementation under the healthy, hunger-free kids Act of 2010. *Fed. Regist* 50151–50170. [PubMed: 27476196]

- Van Der Horst KTA, Crawford D, Roberts R, Brug J, Oenema A, 2008. The school food environment associations with adolescent soft drink and snack consumption. *Am. J. Prev. Med* 35, 217–223. [PubMed: 18617354]
- Van Kann DHH, De Vries SI, Schipperijn J, De Vries NK, Jansen MWJ, Kremers SPJ, 2016. Schoolyard characteristics, physical activity, and sedentary behavior: combining GPS and accelerometry. *J. Sch. Health* 86, 913–921. [PubMed: 27866384]
- Van Sluijs EM, Jones NR, Jones AP, Sharp SJ, Harrison F, Griffin SJ, 2011. School-level correlates of physical activity intensity in 10-year-old children. *Int. J. Pediatr. Obes* 6, e574–e581. [PubMed: 20854106]
- Vine MM, Harrington DW, Butler A, Patte K, Godin K, Leatherdale ST, 2017. Compliance with school nutrition policies in Ontario and Alberta: an assessment of secondary school vending machine data from the COMPASS study. *Can. J. Public Health* 108, e43–e48. [PubMed: 28425898]
- Webber LS, J CC, Rose D, Rice JC, 2007. Development of ACTION! Wellness program for elementary school personnel. *Obesity* (15 Suppl. 1), 48s–56s. [PubMed: 18073341]
- World Health Organization, 2017. Health Promoting School: an effective approach for early action on Ncd risk factors Available at: <https://www.who.int/healthpromotion/publications/health-promotion-school/en/>. Accessed October 1, 2019.
- Zenk SN, Schulz AJ, Mentz G, House JS, Gravlee CC, Miranda PY, Miller P, Kannan S, 2007. Inter-rater and test-retest reliability: methods and results for the neighborhood observational checklist. *Health & Place* 13 (2), 452–465. [PubMed: 16809060]

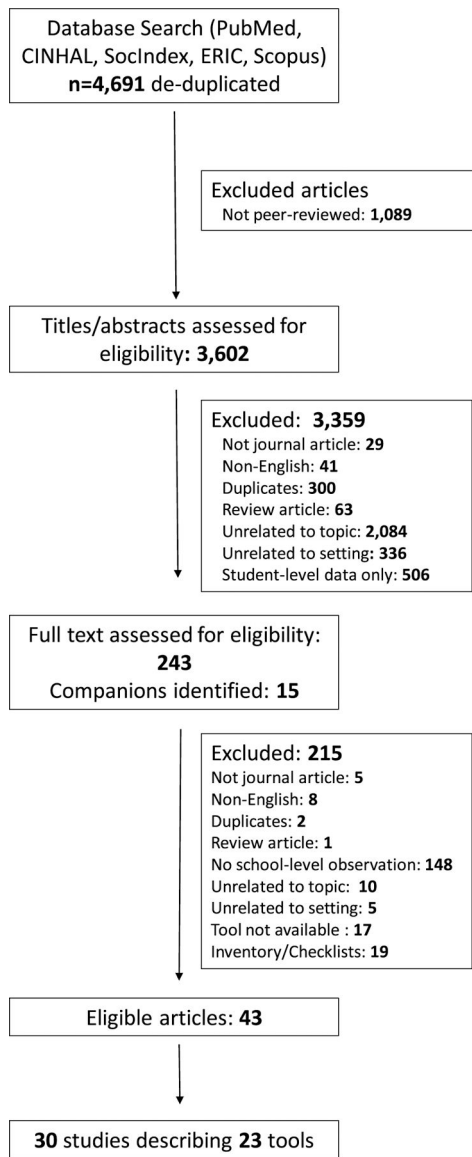


Fig. 1. PRISMA flow diagram for identification, screening, eligibility, and inclusion of studies describing audit tools

Table 1

Description of observational school audit tool studies.

ID	Name of Tool (if named)	Articles	Stated Purpose	Setting	School Type ^c	Item Types ^d	How to Access Tool and Relevant Documents
1	ACTION! Staff Audit	Webber et al. (2007)	Assess physical characteristics of the school and neighborhood as part of formative research for a	New Orleans, Louisiana, USA	ES	Freq. Cat, Open	Tool provided via email from author
2	Adachi et al., 2013	Adachi-Mejia et al. (2013)	Assess and analyze the content, location, and advertising of vending machines in high schools in varying geographic regions	Vermont, New Hampshire, USA	HS	Freq. Cat	Tool described in publication
3	Belansky et al., 2013	Belansky et al. (2013)	Verify whether schools implement planned environment and policy changes following two different intervention	Colorado, USA	ES	Scale, Freq, Cat, Open	Tool provided via email from author
4	Rushing and Asperin (2012)	Rushing and Asperin (2012)	Gather information on branding strategies used in high school nutrition programs	Midwest and Southeast	HS	Freq, Open	Tool described in publication
5	Co-SEA (COMPASS School Environment Application)	Leatherdale et al. (2014a) Hunter et al. (2016); Vine et al. (2017)	Measure aspects of the built environment associated with modifiable determinants of obesity, including using photos to supplement direct observational data	Ontario and Alberta, Canada	HS	Scale, Freq, Cat, Open, Other: photos	Tool and technical report described in publication
5.1	Co-SEA (Adapted)	Godin et al. (2017)	Examine the presence of water sources and industry sponsored sugar-sweetened beverage food kiosks	Guatemala	MS, HS	Scale, Freq, Open, Other:photos	Tool and technical report references provided in publication
6	EAPRS (Environmental Assessment of Public Recreation Spaces)	Colabianchi et al. (2011); Saelens et al., 2006)	Examine correlation between attributes of school playgrounds and playground utilization and activity level	Cleveland, OH, USA	ES	Scale, Freq, Cat, Open	Tool and scoring guide publicly available https://activelivingresearch.org/environmental-assessment-public-recreation-spaces-eaprs-tool
7	ENDORSE school food environment tool	Van Der Horst et al. (2008)	Assess association between eating behaviors and availability of foods and beverages in various areas of the school canteen	Netherlands	MS	Freq, Cat	Tool described in publication
8	Food Decision Environment Tool	Ozturk et al. (2016)	Develop structured tool to assess food-choice architecture and identify behavioral economic intervention points to improve child diet quality	South Carolina, USA	ES	Freq, Cat, Open	Tool described in publication
9	GRF-OT (The Great Recess Framework – Observational Tool)	Massey et al. (2018a, 2018b)	Develop tool to measure contextual factors (e.g., safety, transitions) associated with recess behaviors	Midwestern USA	ES	Scale	Tool, manual/guide, and scoring guide publicly available
10	Hecht et al., 2017	Hecht et al. (2017)	Validate administrator surveys in order to better or more efficiently characterize water access in the schools	San Francisco, USA	ES, MS, HS	Freq, Cat, Open	Tool described in publication
11	ISAT (International Study of Childhood Obesity, Lifestyle and	Katzmarzyk et al. (2013); Broyles et al.	Develop a single tool that can be used world-wide to assess PA environment in schools and its relationship to PA	Australia and 12 other countries	ES	Scale, Freq, Cat, Open	Tool, manual/guide, supporting materials provided in publication

ID	Name of Tool (if named)	Articles	Stated Purpose	Setting	School Type ^c	Item Types ^d	How to Access Tool and Relevant Documents
	Environment School Audit Tool ^b	(2015), Lewis et al. (2016)					
12	Laurie et al. 2017	Laurie et al. (2017)	Assess food garden attributes (size, physical aspects, and use) to improve understanding of sustainable food productive practices	South Africa	ES	Scale, Cat, Open	Tool provided via email from author
13	LCFO (Lunch and Competitive Foods Observation) Form	Ritchie et al. (2015)	Gather information about the nutrition environment (e.g., competitive food venues, school meal items and service, and dining facilities) in schools across the USA	Multiple sites, USA	ES, MS	Scale, Freq, Cat, Open	Tool described in publication, manual/guide provided via email from author
14	PARA (Physical Activity Resource Assessment)	Findholt et al. (2011); Lee et al. (2005)	Identify quantity and quality of publicly accessible PA resources in rural communities	Oregon, USA	ES	Scale, Freq, Cat	Tool and manual/guide described in publication and publicly available https://shaped.fns.usda.gov/library/materials/physical-activity-resource-assessment-para-instrument
14.1	PARA (Adapted)	Pate et al. (2015)	Assess the variety and quality of school ground features and amenities and document evidence of unsociable behaviors	Multiple sites, USA	ES, MS	Scale, Freq, Cat, Open	Tool reference provided and adaptation described in publication
15	Patel et al. 2009	Patel et al. (2009)	Conduct site visits to define state of schools' food environment, and improve environments and policy implementation	Los Angeles, California, USA	MS	Freq, Cat, Open, Other: mapping	Tool provided via email from author
16	School Food Environment Scan	Spruance et al. (2017)	Assess school-level factors (e.g., other food sources, cafeteria environment) related to students' use of salad bars	New Orleans, Louisiana, USA	MS, HS	Scale, Freq, Cat, Open	Tool described in publication, manual/guide provided via email from author
17	School Lunchroom Audits	Thomas et al. (2016)	Provide a process measure to assess fidelity of a Smarter Lunchrooms intervention	Multiple sites, USA	MS	Freq, Open, Other: photos	Tool described in publication
18	SF-EAT (School Food Environment Assessment Tool)	Black et al. (2015)	Assess extent to which schools integrate healthy food initiatives (with equal emphasis on sustainability), and identify potential local and school-specific action	Vancouver, Canada	ES, MS, HS	Scale, Freq, Open	Tool provided via email from author
19	SNDA-III (School Nutrition and Dietary Assessment)	Finkelstein et al. (2008); Briefel et al. (2009); Fox et al. (2009a), Fox et al. (2009b)	Assess USDA school meals and other aspects of the school food environment and policies	Multiple sites, USA	ES, MS, HS	Freq, Cat,	Tool and scoring guide publicly available https://www.fns.usda.gov/school-nutrition-dietary-assessment-study-iii
20	SNEO (School Nutrition Environment Checklist)	Newton et al. (2011)	Assess the cafeteria environment, including both items related to suggested alterations and items related to unhealthy elements	Rural Louisiana, USA	ES, MS	Freq, Cat, Open	Tool provided via email from author
21	SPACE (Spatial Planning and Children's Exercise) Checklist	Van Kann et al. (2016)	Investigate influence of schoolyard characteristics (equipment, green space, upkeep) with child PA levels	Southern Limburg, Netherlands	ES	Scale, Freq, Cat	Tool described and provided in publication
22	SPAN-ET (School Physical Activity and	John et al. (2016)	Measure the school PA/nutrition environment categories related to	Oregon, USA	ES	Scale, Cat, Open	Tool and manual/guide provided via email from author

ID	Name of Tool (if named)	Articles	Stated Purpose	Setting	School Type ^c	Item Types ^d	How to Access Tool and Relevant Documents
	Nutrition Environment Tool)		childhood obesity prevention in order to quantify environmental change				
23	SPEEDY school grounds audit	Jones et al. (2010); Van Sluijs et al. (2011); Mantjes et al. (2012)	Measure the environmental characteristics of schools' external grounds, including their suitability for PA	Norfolk, England	ES	Scale, Freq, Cat	Tool described in publication
23.1	SPEEDY school grounds audit (Adapted)	Harrison et al. (2016)	Measure the environmental characteristics of schools' external grounds, and assess if and how these characteristics change between primary and secondary school	Norfolk, England	MS	Scale, Freq, Cat	Tool reference provided, adaptation described in publication, manual/guide provided via email from author
23.2	SPEEDY school grounds audit (Adapted)	Dias et al. (2017)	Evaluate how quality/quantity of outdoor play structures correlates with practicing PA in physical education class and recess	Passo Fundo, Rio Grande do Sul, Brazil	HS	Scale, Freq	Tool reference provided and adaptation described in publication
23.3	SPEEDY school grounds audit (Adapted)	Tarun et al. (2017)	Measure and describe the characteristics of the outdoor school grounds in order to inform future intervention targets	Delhi, India	ES, MS, HS	Scale, Freq, Cat	Tool reference provided and adaptation described in publication
23.4	SPEEDY school grounds audit (Unadapted)	Chalkley et al. (2018)	Assess opportunities for PA in the school environment prior to program implementation	East Midlands, UK	ES	Scale, Freq, Cat	Tool reference provided in publication
23.5	SPEEDY school grounds audit (Unadapted)	Hyndman and Chancellor (2017)	Describe active play opportunities in secondary school environments in order to inform policy/practice for adolescents	New South Wales, Australia	MS	Scale, Freq, Cat	Tool reference provided in publication

^b ISAT tool was partially adapted from SPEEDY tool.

^c ES = Elementary; MS = Middle; HS=High/Secondary.

^d Scale = Rating scale; Freq = Frequency/Counts; Cat = Categorical or Binary; Open = Open Response.

Table 2
Comprehensiveness (out of 8 possible domains + other) of Observational School Audit Tools (n = 23).

ID	Tool Name (or study author)	Total Domains ^a	Water access (presence, quality of water sources)	Marketing (presence of posters, branding)	Cafeteria (layout, service line)	Vending/A la carte (location, content)	Classroom (n/a)	Indoor Play Area (equipment condition, number of facilities)	Outdoor Play Area (equipment condition, number of facilities)	Garden/Landscape (size, condition of garden)	Other ^b
1	ACTION! Staff Audit	8	X	X	X	X	X	X	X	X	X
2	Adachi et al., 2013	1				X					
3	Belansky et al., 2013	5		X	X	X		X		X	
4	Branding Checklist	2		X	X						
5	Co-SEA	6	X	X	X	X	X	X			
6	EAPRS	1						X			
7	ENDORSE	2			X	X					
8	Food Decision Environment Tool	2		X	X						
9	GRF-OT	1						X			
10	Hecht et al., 2017	1	X								
11	ISAT	6	X			X		X	X	X	X
12	Laurie et al.,	1								X	
13	LCFO	3	X		X	X				X	
14	PARA	2	X					X			
15	Patel et al.,	4			X	X		X	X		
16	School Food Environment Scan	2	X	X							
17	School Lunchroom Audits	2		X	X						
18	SF-EAT	2			X	X					
19	SNDA-III	1				X					
20	SNEO	1			X						
21	SPACE Checklist	1							X		
22	SPAN-ET	5	X		X		X	X	X	X	
23	SPEEDY	4	X					X	X	X	X
	Total Tools		9	7	12	10	0	5	11	6	3

^aMean domains covered = 2.7 (SD = 2.0).
^bOther: active transport infrastructure (ISAT, SPEEDY), facilities (e.g., fast food restaurants) in view (ISAT, ACTION).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Practicality reporting of observational school audit tool studies (n = 30).

ID	ACTION: Staff Audit	How many items in observational component?	How long does it take to complete?	What device is used (i.e., paper vs electronic)?	How many raters are there per school?	What is the expertise of the rater(s)?	How are the rater(s) trained?	Total Items Reported
1	ACTION: Staff Audit	50+			1 to 2	Research staff assisted by school personnel		3
2	Adachi et al, 2013	~4 items for each vending machine						1
3	Belansky et al, 2013	80+	~1 day		2			3
4	Branding Checklist	24			2	Research staff		3
5	Co-SEA			Software application (audit + photos)	1			2
5.1	Co-SEA Unadapted				1 (with occasional auditor)	Research staff	2 in-person trainings	3
6	EAPRS	600+ items in 14 exposure categories	Average 67 min	Pencil/paper	2	Research staff	Trained by senior staff, supervised audit	6
7	ENDORSE	38 items in 4 categories			2			2
8	Food Decision Environment Tool	50+	1–2 h		1 to 2	Economics and public health nutrition faculty		4
9	GRF-OT	7 items in 2 categories	30–45 min (before and during recess)		2	Research staff	Multi-day in-person series, supervised audit	5
10	Hecht et al, 2017				2	Research staff		2
11	ISAT	Range: 21–44 (Countries varied) in 6 categories	57 min on average (Range: 15–160)	Pencil/paper	1 (5% had 2)	Research staff	In-person training, web-based exam, supervised audit	6
12	Laurie et al, 2017	13				National Department of Basic Education fieldworkers		2
13	LCFO	>50	~25 min (less in elementary schools)		1 to 2	Research staff	2.5 h in-person, supervised audit	5
14	PARA	~50	10 min on average (up to 30 min)		2	Research staff	4–8 in-person and 6 field hours, supervised audit	5
14.1	PARA (Adapted)	~40			1 (with occasional auditor)	Community liaisons hired by research staff	In-person, supervised audit	4

ID	How many items in observational component?	How long does it take to complete?	What device is used (i.e., paper vs electronic)?	How many raters are there per school?	What is the expertise of the rater(s)?	How are the rater(s) trained?	Total Items Reported
15	Patel et al., 2009	-25	1.5 h, before and during lunch	At least 2	Community-academic partnership team	3 h in-person training	5
16	School Food Environment Scan	52	1 lunch period (over 2 different days)	Pencil/paper + photos	2	Enumerator/Dietetic Intern	5
17	School Lunchroom Audits	Range: 9-15 (Intervention groups varied)			Research staff		2
18	SF-EAT	8 detailed items + photos		Pencil/paper + photos	1 to 2	Research staff	4
19	SNDA-III	20		Pencil/paper		Research staff	2
20	SNEO	25				Trained by senior staff	3
21	SPACE Checklist	14			2	Research staff	3
22	SPAN-ET	50+ items in 27 categories	10-12 h (includes non-obs. Items)		2	County extension educators, school personnel	5
23	SPEEDY	44 items in 6 categories	30 min	Pencil/paper	2 to 5	Full day in-person, supervised audit	5
23.1	SPEEDY (Adapted; Dias et al., 2017)	8 items in 1 category		Pencil/paper	1	Research staff	4
23.2	SPEEDY (Adapted; Harrison et al., 2016)	47 items in 6 categories		Pencil/paper		Research staff	3
23.3	SPEEDY (Adapted; Tarun et al., 2017)	39 items in 6 categories		Pencil/paper	2 (sometimes accompanied by 3rd school official)	Research staff	5
23.4	SPEEDY (Unadapted; Chalkley et al., 2018)			Pencil/paper	1	Research staff	3
23.5	SPEEDY (Unadapted; Hyndman and Chancellor, 2017)	44 items in 6 categories	30 min	Pencil/paper			3

Table 4

Quality reporting of observational school audit tool studies.

ID	Was formative or pilot testing done, or are adaptations described?	What type (if any) of reliability testing was done? What were the results?	What type (if any) validity testing was done? What were the results?	Is the scoring protocol described?	Total Items Reported
1	ACTION! Staff Audit New; pilot study				1
2	Adachi et al., 2013			Locations described, then other items summed to reflect total number of machines, filled slots, and machine-front advertising per school	1
3	Belansky et al., 2013			Combined with other data to describe implementation changes, then qualitatively classified changes as effective, promising or emerging.	1
4	Branding Checklist New; pilot-tested at first school			Combined with other data sources, qualitatively analyzed using constant comparative method to find patterns	2
5	Co-SEA Adapted from Endorse and SPEEDY; Pilot tested to work through technical difficulties			Scored similarly to ENDORSE and SPEEDY, with change scores calculated from Year 1 to Year 2	2
5.1	Co-SEA Unadapted Used unadapted from original COMPASS study				1
6	EAPRS New; input from parks officials/users and made revisions over several iterations	Interrater (% agreement, ICC, Kappas: 66% items had good-excellent reliability)	Face (Several rounds of input from parks and rec staff and park users)	Variable created for each exposure category (summed binary or frequency items, averaged categorical items)	4
7	ENDORSE New; Reviewed by experts and pilot-tested			Summed or counted, re-coded into 8 “availability” variables, which were dichotomized or categorized into tertiles	2
8	Food Decision Environment Tool New; developed using behavioral economics theories, modified based on feedback from school/study stakeholders throughout	Inter-rater (system to resolve discrepancies during analysis, conducted peer debriefing meetings to clarify)	Trustworthiness of methods (e.g., credibility, transferability, dependability, confirmability) were established and described	Data from observational form were summarized and triangulated using field notes, and analyzed qualitatively for emerging themes	4
9	GRF-OT New; developed over several iterations with input from experts and field testing through PlayWorks	Inter-rater (weighted Kappa: 0.54–1.00, Scale ICC: 0.84); Test-retest (ICC: 0.95)	Convergent (associated with activity levels); Content (fit assessed using exploratory structural equation modeling)	4-category items summed within subdomains	2
10	Hecht et al., 2017	Inter-rater (Kappa: 0.88–1.00)			1
11	ISAT Adapted from SPEEDY and IDEA, then customized by country	Inter-rater (% agreement: 83.9–100%; Kappa: 0.61–0.96)	Construct (could discriminate child PA)	Binary items were reported, and items were also summed within each category	4

ID	Was formative or pilot testing done, or are adaptations described?	What type (if any) of reliability testing was done? What were the results?	What type (if any) validity testing was done? What were the results?	Is the scoring protocol described?	Total Items Reported
12	Laurie et al, 2017 New; developed using pre-existing policies and guidelines and piloted in 9 schools		between highest and lowest quintile schools)	Categorical data were expressed as frequencies and percentages	2
13	LCFO Adapted from SNDA and unpublished tools, based on input from nutrition professionals	Inter-rater (% agreement: >80% with gold standard researcher; monthly quality control review)			2
14	PARA Used unadapted from original PARA study of community physical activity resources	Type not specified (rs > 0.77)		Scored according to original PARA protocol (frequency of features, amenities, incivilities are summed; quality presented as a 3 or 4 item scale)	3
14.1	PARA (Adapted) Adapted from original PARA	Inter-rater (% agreement: >80% with research lead; monthly quality control review)			2
15	Patel et al, 2009 New; developed by members community advisory board in several iterations, including a mock site visit	Inter-rater (Kappa = 0.65–1.0; ICC = 1.0)		Observers compare records to the foods/beverages that align with policy; Other info was qualitatively coded with other data sources to identify themes	3
16	School Food Environment Scan New; developed based on literature review and existing policy documents, then tested for feasibility in 7 schools	Inter-rater was not conducted because observers completed it together		Binary items summed into scale and dichotomized; frequency items summed and dichotomized (some vs none)	2
17	School Lunchroom Audits Adapted from previous iterations of SNDA study		Face (circulated to Co-Is and project partners)	Items are combined with field notes and photos to generate scale score for each service line; summed across all service lines in each school	1
18	SF-EAT New; developed based on literature review and existing policy documents, then tested for feasibility in 7 schools			Items combined with other data sources into 6 pre-determined domains, each scored 1–5 based on extent to which initiatives are happening	3
19	SNDA-III Adapted from previous iterations of SNDA study			Binary variables on audit were combined with other data sources and summed in 3 different categories	2
20	SNEO New; Conducted Q-sort with 8 research staff to select items	Inter-rater (Gewt's AC1 = 0.75); Internal consistency (Cronbach's a = 0.77–0.85)		Two subscales were created: recommended and non-recommended items	3
21	SPACE Checklist Used unadapted from SPACE (Spatial Planning and Children's Exercise) study, but applied to school instruments	Inter-rater (system to resolve discrepancies on-site)			2
22	SPAN-ET Adapted from several existing instruments	Inter-rater (Percent agreement: 80.8–96.8%; Kappa: 0.61–0.94)	Face and content (field tested, with school	Binary items are summed within each category, then categories are explained by a 4-item scale	4

ID	Was formative or pilot testing done, or are adaptations described?	What type (if any) of reliability testing was done? What were the results?	What type (if any) validity testing was done? What were the results?	Is the scoring protocol described?	Total Items Reported
23	New, but based on existing green space instrument	Inter-rater (% agreement: 76–90%; Kappa: 0.67–1)	personnel provided subject-matter expertise Face (draft sent to 3 experts): Construct (could discriminate child PA between highest and lowest quintile schools)	Binary items were summed, frequencies were weighted by response mean, scales were weighted, then all were summed within each category	2
23.1	Adapted from SPEEDY, used only sports and play facility category <i>SPEEDY</i> (Dias et al, 2017)			Scored according to original SPEEDY protocol	2
23.2	Slightly adapted from SPEEDY, added 3 facilities commonly recorded as 'other' in original audit <i>SPEEDY</i> (Harrison et al, 2016)			Scored according to original SPEEDY protocol	3
23.3	Adapted from SPEEDY, removed few items and added "comments" (advised by local experts) <i>SPEEDY</i> (Tarun et al, 2017)	Inter-rater (Kappa scores: 0.4–1.0, % agreement: 61.9%–100.0%)		Scored according to original SPEEDY protocol	3
23.4	<i>SPEEDY</i> (Unadapted; Chalkley et al, 2018)				0
23.5	Used unadapted from original SPEEDY study <i>SPEEDY</i> (Unadapted; Hyndman and Chancellor, 2017)			Scored according to original SPEEDY protocol	2

Table 5

Applicability reporting of observational school audit tool studies.

ID	Does tool relate to a policy, program or ongoing intervention study?	Are tool limitations described?	Are future/potential uses of the tool described?	Are results or analyses reported to schools or districts?	Do authors provide a “call to action” for schools based on findings?	Total Items Reported
1	ACTION! Staff Audit Formative research for ACTION! (group randomized staff wellness study)					1
2	Adachi et al. 2013	Some information cannot be described, including: calorie content, comparing water to flavored water, replacement frequency, or sold out status			Be aware that vending machines are still prevalent, and a potent source of both energy-dense, low-nutrient calories as well as advertising to students	2
3	Belansky et al. 2013	May lack generalizability (only used in rural schools)			Appoint staff champions to coordinate/monitor changes; involve administrator in school-university-partnership, support changes that will improve student behavior	3
4	Branding Checklist	Small sample size			Provides six strategies for schools/districts to develop school nutrition brand personality; school professionals should be trained to implement strategies	2
5	Co-SEA	Measurement error may have occurred when assessing condition of facilities	Photo-taking aspect can inform future methods; Could harmonize data collection across school based studies for generalizability	All schools receive School Health Profiles, which include school-specific data, evidence-based recommendations, and contact information to take action		4
5.1	Co-SEA (Unadapted)	Should have been adapted with unique consideration of Guatemalan school culture	Can use in other countries; recommend adaptation and translation due to school differences across countries	All schools receive School Health Profiles, which include school-specific data, evidence-based recommendations, and contact information to take action	Increase availability of free drinking water, and decrease access, restrict marketing, and enforce sale legislation on sale of sugar-sweetened beverages	4
6	EAPRS	Some confusion with scale items; subjective items had poor reliability; dynamic items not captured; lack variability; may lack generalizability	Could be adapted for parks or playgrounds in other regions to create more user-friendly PA spaces at school;			3

ID	Does tool relate to a policy, program or ongoing intervention study?	Are tool limitations described?	Are future/potential uses of the tool described?	Are results or analyses reported to schools or districts?	Do authors provide a "call to action" for schools based on findings?	Total Items Reported
7	ENDORSE ENDORSE (Environmental Determinants of Obesity in Rotterdam Schoolchildren, prospective 2-year study in adolescents)	Crude constructs of availability; cross-section; possible non-sensitive categorization; no gold standard to test validity	establish predictors of park use			2
8	Food Decision Environment Tool	As a result of tailoring, it may lack generalizability to other settings where choice points may be different	Useful approach to identify targets for research and intervention planning that is aimed toward improving quality of children's food choices			2
9	GRF-OT Aligned with Playworks (nonprofit organization aiming to help schools create recess and play environments for all children)	Framework represents an adult view of recess quality and does not get child perspective; reliability data limited to one large urban public school district.	Could be used to evaluate CDC and SHAPE America recess initiatives nationwide, to ensure consistency and inclusivity of key context variables			3
10	Hecht et al, 2017					0
11	ISAT International Study of Childhood Obesity, Lifestyle, Environment (ISCOLE) (cross-sectional, lifestyle correlates of childhood obesity in 12 countries)	Possible "ceiling effect" with amount of equipment; tool does not record aesthetic qualities; inter-rater only reliability measure	Use of single tool across many countries can facilitate global work to promote healthy school environments		Shift focus from increasing facilities and sports to addressing other drivers of PA (e.g., PE and fitness programs, active breaks and curriculum, etc)	4
12	Laurie et al, 2017 South African National School Nutrition Program's Sustainable Food Production in Schools (SEFPS) pillar	Single day of observation may not fully reflect whether school meets USDA meal standards			Seek outside support to sustain gardens, and links should be created to improve consistency and training, increase garden size, production, equipment, management	2
13	LCFO Healthy Communities Study (longitudinal study of associations between community programs and policies and child/adolescent behaviors and health across USA)					2
14	PARA (Adapted) Healthy Communities Study (longitudinal study of associations between community programs and policies and child/adolescent behaviors and health across USA)					1

ID	Does tool relate to a policy, program or ongoing intervention study?	Are tool limitations described?	Are future/potential uses of the tool described?	Are results or analyses reported to schools or districts?	Do authors provide a "call to action" for schools based on findings?	Total Items Reported
14.1	PARA (unadapted) Mixed methods cross-sectional study to identify environmental influences on children's physical activity and diet in rural communities/schools	Small sample size, only in rural schools	Also used for other neighborhood physical activity resources (e.g., parks, trails, etc)		Consider low-cost ways to make environment health-promoting (e.g., walking school bus, exercise breaks); provide age-appropriate and fun games at recess	4
15	Patel et al., 2009 Community-academic partnership's evaluation of Los Angeles School District's 2005 Cafeteria Improvement Motion (aims to increase NSLP participation, improve school meal marketing)	Qualitative analysis limits quantification of results; small sample size	Other community-academic teams, including school policy stakeholders, should conduct similar site visits to assess policies are implemented as intended	Enabled community-academic partnership to determine priorities for program to help LAUSD translate Cafeteria Improvement Motion into practice	School-specific actions taken as a result of the findings include: creating larger signs to display menu, ordering pre-sliced fruit, offering free, chilled water	5
16	School Food Environment Scan Let's Move Salad Bars to Schools Initiative	May lack generalizability (only used in schools with high concentration of African American students)			Consider promotion and placement to increase salad bar use; Address individual factors that may lead to use	3
17	School Lunchroom Audits National School Lunch Program requirements and Smarter Lunchroom strategies	Single point-in-time audit			To better implement Smarter Lunchrooms, schools need to improve provider engagement and buy-in	3
18	SF-EAT Think&EatGreen @School Alliance (community-university research alliance), provincial guidelines for food and beverage sales in British Columbia, Canada	May lack generalizability (only Vancouver public schools); scoring process not designed to track small differences; cross-sectional	Schools could use as a needs assessment to understand engagement in each domain and/or determine future actions; Could modify for other local or national contexts or priorities using similar development process		School stakeholders should communicate frequently about a clear definition of sustainable food, and develop policies that incentivize and train food service workers to promote sustainable food use	4
19	SNDA-III 2004 Child Nutrition and WIC Reauthorization Act (federal act require schools to have local wellness policies)	Did not investigate the hours when vending machines were accessible to students			Districts should include specific policy language to prohibit certain foods/beverages in all school areas, and work with food and nutrition professionals	3
20	SNEO Louisiana Health (randomized controlled trial to test efficacy of two multi-component, school-based childhood obesity prevention interventions)	Small sample size	Methods can be useful in developing future process evaluation measures, and expanding to include dose and reach			3
21	SPACE Checklist Active Living (study encouraging PA in primary school children in Southern Limburg, Netherlands)	Limited variation was found for schoolyard characteristics			Ensure presence of fixed equipment and opportunities to use that equipment, consider amount/type of green space	3

ID	Does tool relate to a policy, program or ongoing intervention study?	Are tool limitations described?	Are future/potential uses of the tool described?	Are results or analyses reported to schools or districts?	Do authors provide a "call to action" for schools based on findings?	Total Items Reported
22	SPAN-ET Federal policies, best practices, and SNAP-Ed priorities	May lack generalizability (small rural sample); not able to correlate findings with actual child behavior	Could repeat beyond Oregon to assess ability to detect change between schools; could pair with annual behavioral surveillance data to help schools meet national agendas	Tailored results were presented to school stakeholders by extension educators in various formats, including best practice resources specific to "poor" categories	Tailored calls-to-action (and accompanying resources) were provided for each school depending on their scores to incorporate, expand open hours	5
23	SPEEDY SPEEDY (Sport, Physical Activity and Eating Behavior: Environment Determinants in Young People, cohort study to understand patterns of influence on diet/PA among children)	Not sensitive to between-school variability; some subjectivity; hard to differentiate sports; no loose equipment captured; Kappa not good for binary items				3
23.1	SPEEDY (Adapted; Dias et al, 2017)s	Only addresses one factor related to environment (structures), only cross-sectional			Schools should have spaces with good structural conditions to favor planned activities, which can increase student PA	2
23.2	SPEEDY (Adapted; Harrison et al, 2016)	No reliability or validity for middle school, (items may be less relevant and others may be missed); no loose equipment or access policies captured				2
23.3	SPEEDY (Adapted; Tarun et al, 2017)	May lack generalizability (only private schools); no focus on indoor areas; reliability/validity not re-tested for Indian schools			Need multi-pronged policy approach to promote PA, including schools supporting curricula with PA time, investing in and maintaining grounds equipment and space	2
23.4	SPEEDY (Unadapted; Chalkley et al, 2018)	Marathon Kids UK (primary school-based running program delivered by Kids Run Free and Nike)		Protocol indicates that findings will be shared with schools and other stakeholders at a joint dissemination event		3
23.5	SPEEDY (Unadapted; Hyndman and Chancellor, 2017)	No loose equipment captured; school leaders knew of the visit so may have prepared outdoor areas			Schools should address differences between primary and secondary, provide opportunities for social and non-competitive active play, update garden features, facilities, surface type	2