



Article

Adaptation and Validation of the Program Sustainability Assessment Tool (PSAT) for Use in the Elementary School Setting

Alix Hall ^{1,2,3,4,*}, Adam Shoesmith ^{1,2,3,4}, Rachel C. Shelton ⁵, Cassandra Lane ^{1,2,3,4}, Luke Wolfenden ^{1,2,3,4} and Nicole Nathan ^{1,2,3,4}

¹ School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, NSW 2308, Australia; Adam.Shoesmith@health.nsw.gov.au (A.S.); Cassandra.Lane1@health.nsw.gov.au (C.L.); Luke.Wolfenden@health.nsw.gov.au (L.W.); Nicole.Nathan@health.nsw.gov.au (N.N.)

² Hunter New England Population Health, Wallsend, NSW 2287, Australia

³ Hunter Medical Research Institute, New Lambton Heights, NSW 2305, Australia

⁴ Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, NSW 2308, Australia

⁵ Mailman School of Public Health, Columbia University, New York, NY 10032, USA; rs3108@cumc.columbia.edu

* Correspondence: Alix.Ivers@health.nsw.gov.au

Citation: Hall, A.; Shoesmith, A.; Shelton, R.C.; Lane, C.; Wolfenden, L.; Nathan, N. Adaptation and Validation of the Program Sustainability Assessment Tool (PSAT) for Use in the Elementary School Setting. *Int. J. Environ. Res. Public Health* **2021**, *182*, 1414. <https://doi.org/10.3390/ijerph182111414>

Academic Editors: Samantha M. Harden, Scherezade K. Mama, Heather Leach, Laura E. Balis

Received: 22 September 2021
Accepted: 27 October 2021
Published: 29 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: There is a lack of valid and reliable measures of determinants of sustainability specific to public health interventions in the elementary school setting. This study aimed to adapt and evaluate the Program Sustainability Assessment Tool (PSAT) for use in this setting. An expert reference group adapted the PSAT to ensure face validity. Elementary school teachers participating in a multi-component implementation intervention to increase their scheduling of physical activity completed the adapted PSAT. Structural validity was assessed via confirmatory factor analysis. Convergent validity was assessed using linear mixed regression evaluating the associations between scheduling of physical activity and adapted PSAT scores. Cronbach's alpha was used to evaluate internal consistency and intracluster correlation coefficients for interrater reliability. Floor and ceiling effects were also evaluated. Following adaptation and psychometric evaluation, the final measure contained 26 items. Domain Cronbach's alpha ranged from 0.77 to 0.92. Only one domain illustrated acceptable interrater reliability. Evidence for structural validity was mixed and was lacking for convergent validity. There were no floor and ceiling effects. Efforts to adapt and validate the PSAT for the elementary school setting were mixed. Future work to develop and improve measures specific to public health program sustainment that are relevant and psychometrically robust for elementary school settings are needed.

Keywords: sustainability; sustainment; maintenance; physical activity; measurement; development; reliability; school



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Promoting healthy behaviours during childhood is paramount to positive health and wellbeing [1,2]. Schools are an important setting for health promotion activities for children [3]. Numerous studies have investigated the effect of school-based public health interventions on student health outcomes, with systematic review evidence illustrating a positive impact on some outcomes [4–9]. In relation to specific physical activity programs, some reviews have found that formal policies and programs that focus on increasing and improving the quality of physical education [9–11], integrating short bouts of physical activity breaks throughout the day [10], provision of after-school physical activity pro-

grams [9,10], and promoting active transport to school [10,11] may promote physical activity [10,11] or physical fitness in children [9], although findings have been mixed. To optimise the benefits of school-based public health interventions, it is necessary that their delivery is continued over time, especially after external research support for the intervention has been withdrawn [12,13]. This is often referred to as intervention sustainment, which has been defined as the continued delivery of an intervention over time [14,15].

Sustainment of health-related interventions is a common challenge across a variety of settings and populations [12,13,16]. It has been found that once research support has been withdrawn, the full delivery of evidence-based interventions often declines or ceases [12,17–19]. For instance, in a recent review of the sustainment of school-based public health interventions, none of the 18 interventions assessed were sustained in their entirety, with at least one of the intervention components no longer delivered once external funding was withdrawn [18]. Failure to sustain delivery of public health interventions in schools results in a loss of population benefit and wasted resources. It may also impact communities' trust in such programs and influence future uptake or participation [13].

An important part of continuing the delivery of school-based interventions is understanding and addressing the factors that impact intervention sustainment (i.e., determinants of intervention sustainment). It has been argued that multi-level determinants of intervention sustainment differ to some extent across diverse settings and populations [13], indicating the importance of understanding the specific determinants that may be salient within a particular setting (e.g., schools). Knowledge in this area is increasing with the publication of several recent reviews that aimed to identify barriers and facilitators to the sustainment of school-based public health interventions [18,20,21]. While these reviews provide important information on the possible determinants of the sustainment of public health interventions in schools, the factors identified are not always consistent or easily synthesised, due in part to the wide variation in the terminology, methods and measures used to classify determinants of sustainment in individual studies [20–22].

To this end, we must have consistent and comprehensive understanding of the determinants of sustainment. This requires the availability and use of valid, reliable and pragmatic measures that are appropriate and have been validated for the school setting. While there has been an increase in the number of theoretically informed and psychometrically evaluated measures of determinants of sustainment [23,24], few have been designed and evaluated for use specifically in the school environment. For example, in a recent review of implementation-related measures for school-based health policies, only three measures addressed sustainability [22].

In another review of implementation measures [24], only one of the sustainability measures was designed, developed and evaluated specifically for a school environment (the "School-Wide Universal Behavior Sustainability Index-School Teams" scale or SUBSIST). While the SUBSIST questionnaire is specific to the school context and has scored highly on psychometric and pragmatic rating scales in previous reviews [23,24], it was designed for a specific intervention [25] the "School-Wide Positive Behaviour Support" [25]. Consequently, SUBSIST may not be appropriate to identify factors impacting on the sustainment of school-based programs more broadly. If we are to build a comprehensive understanding of what factors need to be addressed in order to support the sustainment of school-based public health interventions, we must develop valid and reliable measures that are generalizable and relevant to a broad range of interventions. Such endeavours will also help to improve replication of findings, which is essential for building high-quality evidence.

There are, however, a number of other, more general measures of determinants of sustainment that may be appropriate for adaptation to the school setting [15,22–24]. In particular, the Program Sustainability Assessment Tool (PSAT), created by researchers at Washington University, is a 40-item tool that assesses the capacity of public health programs to be sustained across eight key domains, including: Environmental Support, Fund-

ing Stability, Partnerships, Organizational Capacity, Program Evaluation, Program Adaptation, Communications and Strategic Planning [26]. The measure has illustrated some evidence of reliability and validity in the context of chronic disease prevention programs [26]. Its development was informed by a sustainability-specific framework [27], which may reflect a range of important determinants of sustainment of public health interventions generally. Additionally, the measure is relatively short, easy to use [26] and is highly accessible and flexible with a license that allows for adaptations to be made [28]. Furthermore, the PSAT is targeted towards assessment of public health programs across a broad range of settings, including education [28]. While the current items and domains of the PSAT may require some amendments to allow for more specific assessment of sustainability determinants of public health programs in the elementary school setting, the generalised nature of the PSAT, with its solid theoretical underpinnings, strong pragmatic qualities and emerging psychometric properties, makes it a strong candidate for adaptation into a measure that is more specific to the school setting.

The overall objective of this study was to adapt and evaluate the relevance, reliability and validity of an adapted version of the PSAT to assess a physical activity policy within the elementary school setting. The specific aims were to:

- a. Adapt the PSAT domains and items to reflect the determinants of sustainment of a physical activity policy in the elementary school setting. In this specific study a physical activity policy was the target public health program, but with the view that the measure could be extended in future efforts to cover other public health programs in this setting.
- b. Evaluate the following psychometric properties of the adapted PSAT: structural validity, convergent validity, internal reliability, interrater reliability, floor and ceiling effects and norms.

2. Materials and Methods

2.1. Phase 1: Face and Content Validity

Content validity is the extent to which the items represent the constructs that a scale is designed to measure [29,30]. Face validity is a component of content validity and relates to the degree to which end-users determine the items as being an appropriate representation of the target constructs [29]. A group of experts and members of the target population, consisting of implementation scientists ($n = 8$), elementary school teachers ($n = 4$) and public health service delivery specialists ($n = 4$), reviewed and adapted the PSAT to ensure that the items adequately reflected the definitions of the domains, and were relevant and acceptable to the elementary school setting.

2.2. Phase 2: Psychometric Evaluation

Following adaptation of the PSAT, we undertook a formal psychometric evaluation of the adapted measure. The methods used are described below.

2.3. Design

The reliability and validity of the adapted PSAT were assessed as a secondary analysis using a convenience sample. Data were obtained from cross-sectional data collected from teachers following receipt of active implementation support from the intervention arms of two school-based cluster randomised controlled trials. Both trials assessed the impact of similar multi-strategy implementation strategies on increasing the mean minutes of physical activity scheduled by teachers across the school week. The first study was a pilot trial [31], with six schools randomised to receive the implementation strategies. Sustainment data collection occurred 18-months following active implementation support. The second study was a larger effectiveness trial [32], with 31 schools randomised to receive the implementation strategies. Sustainment data collection occurred six months

after active implementation support ceased. Despite slight differences in the implementation strategies being assessed in the two trials, the physical activity intervention (i.e., scheduling of weekly classroom physical activity) that was the focus of the adapted PSAT was the same in both trials. Both trials received Human Research Ethics approval from Hunter New England (no. 06/07/26/4.04), the University of Newcastle (no. H-2008-0343) and relevant elementary school bodies.

2.4. Sample and Procedures

Elementary schools from the Hunter New England (HNE) region of New South Wales (NSW), Australia were eligible to participate in the trials if they were not participating in another physical activity intervention, and did not enrol only students who require specialist care [31,32]. HNE is a demographically and geographically diverse region in NSW Australia. It covers an area of approximately 130,000 km², which includes densely populated regions such as metropolitan and regional hubs, as well as more geographically isolated areas such as rural and remote locations [33]. Socioeconomically, HNE is also very diverse, with areas of high wealth and other areas of poverty [33].

Following principal consent, schools were randomised to receive the multi-strategy implementation intervention or usual practice control. All classroom teachers from participating schools were invited to complete a self-report survey independently at three time-points (baseline, follow-up, and sustainment). Completion of the teacher survey was deemed consent to take part. Teachers from intervention schools completed a survey at all three time-points with the adapted PSAT items included in the sustainment data collection time-point. Only data from the sustainment data collection time-point from the intervention schools were included in this study. A complete description of the pilot [31] and effectiveness trials [32] are reported elsewhere.

2.5. Measures

2.5.1. The Adapted Program Sustainability Assessment Tool (PSAT)

Based on stakeholder and end-user feedback (from Phase 1), an adapted version of the PSAT was evaluated. The adapted PSAT was reduced from 40 items across eight domains to 30 items across the following seven domains: Environmental Support, Funding Stability, Organizational Capacity, Program Evaluation, Program Adaptation, Communications and Strategic Planning. The Partnerships domain (5 items) was removed in addition to 11 items across the remaining domains (excluding Environmental Support), as these items were viewed as lacking relevance to this context. There were six additional items created relating to Organisational Capacity (e.g., resources and infrastructure, staff training); and Funding Stability (e.g., process for attending professional development), to ensure that constructs within these domains were adequately covered in relation to the school context. For each item respondents indicated the extent to which they agreed with the item, using a seven-point Likert scale, with response options ranging from “strongly disagree” to “strongly agree”. Consistent with the original measure, all items were positively worded with a higher level of agreement reflecting a greater capacity for program sustainment.

2.5.2. Sociodemographic Characteristics

Teachers were asked to report their sex, year of birth, years of teaching experience, grade level taught and if they were a specialist physical education (PE) teacher.

2.5.3. Sustainment of Scheduled Physical Activity at School

The implementation intervention was designed to increase teachers' scheduling of physical activity across the school week. To assess this outcome teachers were asked to complete a daily activity logbook documenting the number of minutes they scheduled physical activity each day for one school week (5 days). The logbook included the time

and occasion physical activity was scheduled for PE, sport or other structured activities (i.e., energisers and active lessons). This data provided an indication of teachers' implementation of the current state Department of Education policy, which mandates that teachers schedule a minimum of 150 min of physical activity across the school week [34], and which was the target behaviour of the implementation intervention being tested in both trials. Sustainment of this behaviour was considered at 18 months following completion of the implementation intervention in the pilot trial [31] and 6 months in the effectiveness trial [32].

2.6. Statistical Analysis

Statistical analyses were undertaken in SAS version 9.3 and R version 4.0.2. The PSAT was developed based on a reflective measurement model, where it is theorised that the items in the scale are a manifestation of the same underlying construct, and thus are expected to be highly correlated and interchangeable [35]. Consequently, the psychometric indicators assessed in this study assume a reflective measurement model. Prior to assessing any psychometric properties, missing responses and response patterns were assessed for each item to identify any items that were poorly responded to and reviewed for possible exclusion. The polychoric correlations between all pairs of items were also calculated and reviewed to help identify any possible redundancies in the items. Items with a polychoric correlation coefficient above 0.8 [36] were reviewed for possible exclusion. A psychometric evaluation was then conducted on the resulting scale. An overview of the specific psychometric properties and the statistical analyses used are described below.

2.6.1. Structural Validity

Structural validity is the extent to which the items in a scale are an adequate reflection of the hypothesized dimensionality of the construct being measured [35]. Structural validity is a component of construct validity [35], and is often assessed via factor analysis. As the dimensionality of the PSAT has been previously established and thus we have a clear hypothesis of how the items of the scale should relate to one another [37], a confirmatory factor analysis (CFA) proposing a seven-factor structure was conducted. To account for clustering of teachers within schools, we employed the procedures proposed by Hox [38] and outlined by Huang [39] to estimate a level one CFA model using the pooled within-cluster covariance matrix. This analysis provides an unbiased estimate of the model parameters by removing the group-level effects [39]. Pairwise deletion was used in the calculation of the covariance matrix, using all available data. Maximum likelihood was used as the estimation method. Parameter estimates were standardized with variances fixed at one [36]. An initial model assuming no correlation between factors was estimated and then revised to allow for such correlations, as it is reasonable to assume a relationship exists between the theoretical constructs. The following fit statistics and recommended criteria were used to assess the overall adequacy of the model:

- Standardized Root Square Residual (SRMR) < 0.08 [35,40];
- Comparative Fit Index (CFI) > 0.95 [35,40];
- Root Mean Square Error of Approximation (RMSEA) < 0.06 [35,40];
- Model Chi-Squared p -value > 0.05 [40].

To reduce selection bias we pre-specified the above fit indices, selecting those that were recommended as they have been found to be most insensitive to the sample size, model misspecification and parameter estimates used [40]. Modification indices and factor loadings were examined and used to revise the CFA model to ensure the most parsimonious, adequately fitting and theoretically justifiable model was selected. Specifically, items with low factor loadings (<0.40) or cross-loadings were examined for removal or model amendments. Standardized factor loadings and their associated standard error and p -value are reported.

2.6.2. Convergent Validity via Hypothesis Testing

Convergent validity is a component of construct validity of a self-report measure and involves assessing the relationship between the proposed scale and similar constructs [29]. The PSAT is designed to assess the capacity for program sustainability. Thus, if the adapted PSAT accurately reflects its intended construct, i.e., capacity for program sustainability, adapted PSAT scores should be positively related to sustainment of the target behaviour, as has been qualitatively observed to some extent in a previous study using the original PSAT measure [41]. Based on the theoretical construct that the PSAT is intended to measure and findings from previous research, we hypothesised that the seven domains of the adapted PSAT as well as the total adapted PSAT score would be positively related to the teacher's total minutes of scheduled physical activity across the school week at the sustainment data collection phase, as this was the target behaviour that we were attempting to sustain in the two intervention trials.

To assess this hypothesis, adapted PSAT domain scores were calculated by summing together each item in a domain and dividing by the number of non-missing items. Scores were only calculated for teachers who answered a minimum of 50% of items from each domain. Linear mixed regression models were used to assess the relationship between the seven adapted PSAT domains and total adapted PSAT score with teachers' total minutes of scheduled physical activity at the sustainment phase of data collection. A separate model was conducted for each domain and the total score, with a random level intercept for school to account for clustering. A positive relationship between the adapted PSAT scores and the number of minutes teachers scheduled physical activity (i.e., intervention target) was hypothesised.

2.6.3. Internal Reliability

Internal reliability is the extent to which items in a domain or scale are correlated [30]. Cronbach's alpha were calculated for each domain, with values between 0.7 and 0.95 considered acceptable [30].

2.6.4. Interrater Reliability

To assess the degree to which teachers from the same school were reporting similar adapted PSAT scores, we calculated the intraclass correlation coefficients (ICC), ICC(1) and ICC(2), for each domain as well as for the overall adapted PSAT score using a linear mixed effects model. The ICC(1) presents the proportion of variance in the adapted PSAT scores that is attributable to school membership (i.e., between-group) [42]. ICC(2) represents the reliability of the group-adapted PSAT scores [42]. Similar to other psychometric evaluations of implementation measures, a threshold of 0.70 for ICC(2) was used to indicate adequate group level reliability [43].

2.6.5. Floor and Ceiling Effects

Potential responsiveness was evaluated by assessing absence of floor and ceiling effects. Scales that illustrate limited floor and ceiling effects have an ability to capture future changes in the construct being measured. The percentage of respondents reporting the lowest and highest possible score for each domain were calculated. Domains where >15% of respondents obtained the lowest (floor) or highest (ceiling) score were considered indicative of floor and ceiling effects [30].

2.6.6. Norms

The mean, standard deviation, median, minimum and maximum for each of the adapted PSAT domains and total adapted PSAT score were calculated.

3. Results

3.1. Phase 1: Face and Content Validity

Substantial adaptation was required to ensure relevance and face validity of the PSAT for the elementary school setting. Each domain was assessed to ensure its relevance to the school setting with one of the original eight PSAT domains removed: the Partnership domain (5 items). As many health programs and policies relevant to schools are often determined by educational departments or boards, the establishment of partnership with stakeholders was considered outside the scope of individual schools, particularly teachers, and more relevant to the governing bodies of schools. The remaining seven domains (originally consisting of 35 items) were kept and the items within each reviewed and assessed for relevance, appropriateness and acceptability to the school setting. As a result of this process, 11 items from the original PSAT were removed: three from Funding Stability; two from each of the Strategy Planning, Program Adaptation and Communications domains; and one each from the Organisational Capacity and Program Evaluation domains. Most of the items removed were perceived by stakeholders and experts as not relevant to individual schools and more relevant to a school's governing body (e.g., a department of education, school board) (see Table S1). Six additional items were created as they were identified as potentially important determinants of program sustainment in the school setting by the expert group. Specifically, four were added to the Organisational Capacity domain and two to the Funding Stability domain (see Table S1). The 24 remaining original items from the PSAT were reworded to ensure they were relevant and specific to the school setting. For example, wording changes consisted of swapping "the program" as the subject to "my school", as it did not make sense that the program performed the actions in most of the items. We also included the specific program of interest, in our case the scheduling of physical activity, but with a view that this could be changed depending on the specific program being evaluated. This was done to improve clarity and ensure that respondents were interpreting all items consistently and with reference to the program of interest. Finally, we also added in an example to some of the items, to again improve clarity and consistency of reporting (see Table S1 for comparison of the original PSAT items with the amended items). From this process, a total of 30 items across seven domains were included in the teacher survey as part of the initial adapted PSAT and included in the psychometric evaluation.

3.2. Phase 2: Psychometric Evaluation

Two hundred and sixty-one teachers from 33 (89%) schools returned a survey, of which 45 (17%) missed all 30 items and two missed more than half of all items. This left a sample of 214 participants from 30 schools who were included in the quantitative analyses as they answered a minimum of 50% of the adapted PSAT items; 187 participants answered all adapted PSAT items. School and participant characteristics are outlined in Table 1.

Table 1. School and teacher characteristics.

Characteristics	<i>n</i> (%)
School level <i>n</i> = 30	
Sector	
Catholic	14 (47%)
Government	16 (53%)
Region	
Inner/outer regional	9 (30%)
Major city	21 (70%)
Trial	
Pilot	5 (15%)
Effectiveness	25 (85%)

Number of teachers (mean (SD))	7 (5)
Teacher level $n = 214$	
Gender	
Male	33 (15%)
Female	181 (85%)
Employment status	
Permanent full-time	127 (59%)
Temporary full-time	59 (28%)
Permanent part-time	20 (9.4%)
Temporary part-time	8 (3.7%)
Job share *	
Yes	42 (20%)
No	168 (80%)
Trial	
Pilot	31 (16%)
Effectiveness	183 (86%)
Age—mean (SD)	40.7 (11)
Years of teaching—mean (SD)	15.2 (11)

* Frequency and percentages may not equal total sample due to missing values.

3.3. Item Assessment

Missing values were low for all 30 items, ranging from 0.47% to 4.70% (see Table 2). The full range of response options were used for 27 of the 30 items, although a left-hand skew was observed for all 30 items, with less than 3% of respondents utilising the lower end of the response scale, and most participants answering towards the positive end of the scale. For three items from the Organisational Capacity domain, the “strongly disagree” response option was not endorsed by any of the respondents. These items included: “My school has enough trained school champions to support the scheduling of physical activity,” “School champions and teachers at my school have enough supervision and support to implement the scheduling of physical activity” and “The level of school champion/teacher turnover is manageable to sustain the scheduling of physical activity.”

Polychoric correlation coefficients ranged from 0.28 to 0.87. Ten pairs of items recorded polychoric correlations above 0.8, reflecting possible redundancies. Of these 10 items, two were deemed appropriate to remove as they were considered adequately reflected by other items in the scale. The items removed included: “The scheduling of physical activity is well integrated into the operations of our school” from the Organisational Capacity domain, and “Evaluation results of the scheduling of physical activity are used to demonstrate success to funders and other key stakeholders (e.g., P&C, wider school community, etc.)” from the Program Evaluation domain. This resulted in 28 items being included in the psychometric evaluation.

Table 2. Item-level information for the final adapted PSAT.

Domain and Items	Missing N (%)	Standardised Factor Loading (se)	p -Value
Domain: Strategic planning			
My school has a sustainability plan (e.g., to continue the scheduling of the recommended minutes of physical activity long-term).	10 (4.7)	0.74 (0.04)	<0.001
My school’s goals to maintain the scheduling of PA are understood by all stakeholders (e.g., teachers, school champions, principals).	8 (3.74)	0.92 (0.02)	<0.001
My school clearly outlines roles and responsibilities to schedule PA for all stakeholders (e.g., teachers, school champions, principals).	8 (3.74)	0.91 (0.02)	<0.001
Domain: Environmental support			
There are champions within the school advocating for the scheduling of PA (a champion is someone who supports and advocates the policy, this may be your school executive or a teacher within the school).	2 (0.93)	0.70 (0.05)	<0.001

There are champions within the school with the ability to get resources for the scheduling of PA.	2 (0.93)	0.74 (0.04)	<0.001
My school has support from within the broader organisation i.e., DoE/CSO for the scheduling of PA.	4 (1.87)	0.71 (0.04)	<0.001
My school has support from outside our education department/office to help the scheduling of PA.	5 (2.34)	0.63 (0.05)	<0.001
The scheduling of PA for students at my school has strong public and community support.	5 (2.34)	0.68 (0.05)	<0.001
Domain: Program adaptation			
My school adapts or changes the scheduling of physical activity each week as needed (e.g., if PE equipment is damaged and cannot be used, heat wave etc).	2 (0.93)	0.74 (0.04)	<0.001
My school has a process to proactively adapt the scheduling of PA to meet changes in needs of the school community (e.g., to include other school programs).	2 (0.93)	0.82 (0.04)	<0.001
My school makes decisions about which physical activity components are ineffective and should not continue when scheduling PA (e.g., energizers, GoNoodle, running etc.)	3 (1.40)	0.59 (0.06)	<0.001
Domain: Organisational capacity			
School systems (e.g., space, time allocation) are in place to support the scheduling of PA.	1 (0.47)	0.71 (0.04)	<0.001
There are adequate resources and infrastructure within the school to schedule PA.	1 (0.47)	0.69 (0.04)	<0.001
School executives manage staff and other resources effectively to ensure that the scheduling of PA is met.	1 (0.47)	0.83 (0.03)	<0.001
My school has enough trained school champions to support the scheduling of PA.	1 (0.47)	0.74 (0.04)	<0.001
School champions and teachers at my school have enough supervision and support to implement the scheduling PA.	1 (0.47)	0.72 (0.04)	<0.001
<i>Note: The item "the scheduling of PA is well integrated into the operations of our school" was removed during the item assessment process. Items "The level of school champion/teacher turnover is manageable to sustain the scheduling of PA" and "My school has a system for training new school champions/teachers to schedule PA" were removed based on modification indices from CFA.</i>			
Domain: Communications			
My school has communication strategies in place to secure and maintain our school communities' support for scheduling PA.	2 (0.93)	0.75 (0.04)	<0.001
Staff members at my school communicate the need for scheduling PA to the community (e.g., parents)	3 (1.40)	0.84 (0.03)	<0.001
My schools' scheduling of PA increases community awareness of the need for PA in children	4 (1.87)	0.83 (0.03)	<0.001
Domain: Program evaluation			
My school has a system in place to actively evaluate the scheduling of PA (e.g., Improvements in children's PA, student on-task behaviour etc.)	1 (0.47)	0.86 (0.03)	<0.001
My school reports the outcomes of scheduling the recommended minutes of PA (e.g., Improvement in student physical activity levels)	1 (0.47)	0.86 (0.03)	<0.001
Evaluation results inform the planning and implementation of the scheduling of PA.	1 (0.47)	0.81 (0.03)	<0.001
<i>Note: The item "Evaluation results of the scheduling of PA are used to demonstrate success to funders and other key stakeholders (e.g. P&C, wider school community, etc.)" was removed during the item assessment process.</i>			
Domain: Funding stability			
The school takes action to ensure there are ongoing funds to support the scheduling of PA. (e.g., included in annual school budget, funding from P&C)	2 (0.93)	0.75 (0.04)	<0.001
My school has a process in place to allow staff to attend professional development on scheduling PA (i.e., funding for ongoing professional development)	4 (1.87)	0.75 (0.04)	<0.001
My school provides time at work for staff to plan their schedule for meeting the recommended minutes of PA.	3 (1.40)	0.64 (0.05)	<0.001
My school can access a variety of funding sources to help schedule PA.	2 (0.93)	0.81 (0.04)	<0.001

3.4. Structural Validity

The initial model assuming no correlation between factors was a poor fit to the data across all fit indices and was subsequently improved in the first revised model by allowing factors to be correlated (see Table 3). The first revised model met the pre-specified criteria for adequate model fit according to the SRMR fit index but no others. One item from the Organisational Capacity domain was removed based on modification indices

and review due to cross-loading with the Program Evaluation domain (“My school has a system for training new school champions/teachers to schedule PA”) (see Table S1). The second revised model illustrated a slightly better fit to the data, illustrated by the improved fit indices and lower Akaike Information Criteria (AIC) (see Table 3). However, again only the SRMR index met the pre-specified criteria. A third revised model was calculated, further removing another item from the Organisational Capacity domain (“The level of school champion/teacher turnover is manageable to sustain the scheduling of PA”) due to correlation with other items from the Strategic Planning scale (see Table 3). All remaining items were considered theoretically important, and no further amendments were made to the model. The final items with their factor loadings are presented in Table 2.

Table 3. Model fit statistics from confirmatory factor analysis ($n = 214$).

Model	SRMR	CFI	RMSEA	χ^2 (df), p -Value	AIC
Initial model—assuming there is no correlation between factors	0.319	0.644	0.135 (0.128, 0.142)	1517.64 (350), $p < 0.001$	13,259.10
Revised model 1—allowing factors to be correlated	0.073 *	0.803	0.103 (0.096, 0.111)	974.41 (329), $p < 0.001$	12,757.87
Revised Model 2—removing item from Organisational Capacity domain	0.070 *	0.820	0.100 (0.093, 0.108)	864.31 (303), $p < 0.001$	12,248.90
Revised model 3—removing item from Organisational Capacity domain	0.070*	0.823	0.101 (0.093, 0.110)	804.08 (278), $p < 0.001$	11,844.87

* Indices meet the specified accepted criteria.

3.5. Convergent Validity via Hypothesis Testing

Evidence for convergent validity tested via hypothesis testing was lacking, with small, non-significant associations recorded for all seven adapted PSAT domains and the total adapted PSAT score with teachers’ scheduling of weekly minutes of physical activity (Table 4).

Table 4. Domain-level results assessing the internal reliability, floor and ceiling effects, norms, hypothesis testing and interrater reliability.

Domain	Standardised Alpha	Floor n (%) ^a	Ceiling n (%)	Mean (SD)	Median (Q1, Q3)	Minimum and Maximum Score	Hypothesis Testing Coefficient, p -value *	ICC1	ICC2
Strategic planning	0.90	2 (1.0%)	12 (5.9%)	5.05 (1.15)	5.00 (4.33, 6.00)	1.0 and 7.0	-0.06, $p = 0.981$	0.21	0.60
Environmental support	0.84	0	5 (2.4%)	5.12 (0.91)	5.20 (4.60, 5.80)	2.0 and 7.0	1.97, $p = 0.525$	0.22	0.61
Program adaptation	0.77	0	10 (4.7%)	5.17 (1.02)	5.33 (4.67, 6.00)	2.0 and 7.0	2.47, $p = 0.329$	0.10	0.39
Organisational capacity	0.87	0	9 (4.2%)	5.30 (0.93)	5.40 (4.80, 6.00)	2.4 and 7.0	-0.04, $p = 0.99$	0.21	0.59
Communications	0.89	1 (0.5%)	7 (3.3%)	4.79 (1.03)	4.67 (4.00, 5.50)	1.0 and 7.0	0.08, $p = 0.978$	0.21	0.60
Program evaluation	0.92	1 (0.5%)	3 (1.4%)	4.46 (1.12)	4.33 (4.00, 5.33)	1.0 and 7.0	-2.21, $p = 0.376$	0.26	0.65
Funding stability	0.86	1 (0.5%)	7 (3.3%)	4.85 (0.98)	4.88 (4.25, 5.50)	1.0 and 7.0	-1.26, $p = 0.667$	0.37	0.75
Total PSAT	0.95	0	3 (1.4%)	5.00 (0.80)	4.98 (4.46, 5.54)	2.2 and 7.0	0.39, $p = 0.914$	0.30	0.69

^a Percentages may not correspond to total sample due to missing data. * Only those with valid physical activity scheduling data were included in the analysis.

3.6. Internal Reliability

All domain Cronbach's alpha values were between the pre-specified threshold of 0.70 and 0.95, ranging from 0.77 to 0.92, and the total score was 0.95 (see Table 4).

3.7. Interrater Reliability

ICC(1) values ranged between 0.10 and 0.37 for the PSAT domains, indicating that between 10% to 37% of the total variance in adapted PSAT domain scores was attributable to differences between schools. The ICC(1) for the total adapted PSAT scores was 0.30 (Table 4). ICC(2) values ranged from 0.39 to 0.75 for the adapted PSAT domains, with only one domain, the Funding Stability domain, exceeding the 0.70 criteria for acceptable group-level reliability. The ICC(2) value for the total adapted PSAT score was 0.69 (Table 4).

3.8. Floor and Ceiling Effects

Fewer than 15% of respondents obtained the lowest or highest possible score for each of the seven domains. This indicates limited floor and ceiling effects of the measure (Table 4).

3.9. Norms

Domain scores ranged from a mean of 4.46 (SD = 1.12) to 5.30 (SD = 0.93), and a median 4.33 (Q1 = 4.00, Q3 = 5.33) to 5.40 (Q1 = 4.80, Q3 = 6.00), out of a possible range of one to seven (see Table 4).

4. Discussion

This study aimed to adapt and assess the validity and reliability of the PSAT as a measure for assessing the capacity for sustainment of teacher delivery of a physical activity policy in elementary schools. The final adapted 26-item PSAT for the elementary school setting illustrated adequate internal reliability for all domains, with Cronbach's alpha values meeting acceptable thresholds. This finding is consistent with the reliability findings for the original PSAT [26]. However, the criteria for interrater reliability were only met for one domain, the Funding Stability domain, suggesting limited consistency in ratings from teachers from the same school for all other domains. This suggests that different individuals from the same school may have different views on the factors impacting on a school's sustainment of a physical activity policy. Findings relating to indicators of validity were mixed, with inconclusive evidence for structural validity, no evidence of convergent validity from hypothesis testing, and no evidence of floor and ceiling effects, which suggests the potential responsiveness of the measure. The lack of strong evidence for validity suggests it is possible that we are failing to adequately capture the determinants of capacity for program sustainment in a school-based setting, at least in relation to the sustainment of scheduling physical activity across the school week.

Informed by stakeholders, implementation and school expertise, the original PSAT underwent extensive adaptation to the items and domains in order to ensure its relevance and appropriateness to the elementary school setting. This resulted in the entire "Partnerships" domain being removed as it focuses on issues of cultivating connections between the target program and stakeholders [26], an issue that is more relevant to a school's governing body (e.g., educational department or schoolboard) than the individual schools, particularly teachers themselves. A further 11 individual items across six of the remaining seven domains were removed, also due to perceived lack of applicability to the individual school level. An additional six items were created: two to the "Funding Stability" domain and four to the "Organisational Capacity" domain. These new items were included as they were identified by the expert panel as important to the sustainment of health programs in schools (see Table S1 for the full list of amendments made). Furthermore, we found that most of the items had poor coverage of the lower end of the constructs being measured, with little variation and response at the lower end of the scale. These findings suggest that the adapted PSAT may not adequately measure the entire set of underlying

constructs in this context and may need further refinement and specification to ensure it adequately captures both low and high levels of the domains. This is consistent with some previous studies that have used the PSAT and have reported domain scores skewed towards the higher end of the scale [44,45]. However, other studies have reported greater variation across domain scores of the PSAT [41,46–48].

In addition, no statistically significant relationship between the seven adapted PSAT domains or the total adapted PSAT score and teachers' scheduling of physical activity were found. This is of concern, as sustained teacher scheduling of physical activity was the focus and primary outcome of the public health program that our implementation intervention was aiming to continue. Despite our intervention illustrating a significant and maintained effect on this outcome [32], the adapted PSAT failed to differentiate between teachers from the intervention schools with high and low scheduling behaviour, which we propose theoretically should reflect program sustainment. This finding highlights the potential usefulness of the adapted PSAT in its current form as a measure of determinants of sustainment of physical activity programs in elementary schools. However, while we did expect that higher scheduling of physical activity should reflect sustainment, this is only a proxy measure of sustained delivery of the policy and may not be truly reflective of this outcome. Given the multi-dimensional nature of sustainment, future research should ideally measure multiple indicators of sustainment, not just one. Furthermore, evidence of structural validity was limited, with only one of the fit indices from the CFA meeting the pre-specified criteria. Again, this indicates potential limitations in the validity of the adapted PSAT measure. However, most factor loadings were high (>0.40) and the findings from our CFA were only slightly poorer than those obtained from the psychometric evaluation of the original PSAT, where similarly only the SRMR met the criteria we used to determine adequate model fit [26]. Future studies with larger sample sizes and improved CFA methods are required to explore, in greater detail, the complex structure of this measure in order to gain a more robust understanding of the structural validity of the adapted PSAT.

The interrater reliability was only acceptable for the Funding Stability domain, highlighting the lack of consistency in individual teachers' views from the same school on the determinants of capacity for the sustainability of a physical activity policy. It is possible that the specific factors being assessed by the adapted PSAT were not all relevant or appropriate for teachers to answer. While classroom teachers are frontline implementers who are instrumental to the day-to-day delivery of school-based public health programs, they often do not have authority over the organisational and external factors that are a large focus of the PSAT items. While a number of items not relevant to the school setting, or more specifically teachers, were removed to ensure relevance of the adapted PSAT, many of the remaining items are still possibly difficult for teachers to reliably respond to, such as allocation of funding and engagement of external providers to support implementation. Consequently, correct responses to some of the adapted PSAT items may require in-depth knowledge of schools' organisational practices, which principals may decide but individual teachers are unlikely to be aware of. It has been recently argued that many existing measures of sustainment determinants may not be suitable for frontline individuals or practitioners [15], which is possibly the case in this instance. Sustainment is a complex, multidimensional process that is influenced by interactions across a range of multi-level determinants [13]. To completely understand the full range of factors that impact on program sustainment, different information may need to be collected from multiple sources and types of end-users and stakeholders. While it is encouraged that a range of stakeholders complete the PSAT [49], in the context of the school environment it may be important to develop separate measures that cover determinants of sustainment that are relevant to teachers and to school leadership or administrators. Future work is needed to determine the best approach to obtaining a comprehensive and accurate understanding of the determinants of sustainment.

Limitations

There are several limitations that must be acknowledged when interpreting the current study results. First, despite involving members of the target population and experts in the content area to assist in the adaptation of the PSAT, we employed a relatively informal process of adaptation with experts who were involved in both the refinement of the scale and assessment of the face and content validity. If time and resources permit, future measure development studies should strive to include members of the target population in the development of the measure, while conducting cognitive interviews with a separate sample of the target population to ensure face and content validity [29]. Second, while we undertook methods to address the clustered nature of our data [38,39] and used the most conservative model, the method used assumes that the factor structure is the same at the individual and school level [38,39], which may not be the case. Unfortunately, we did not have an adequate sample or fit to explore the possible multilevel structure of the measure and were also limited in the modelling methods we could employ. Furthermore, no hypotheses concerning the multilevel structure of the adapted PSAT have been proposed, further limiting the ability to explore the higher-order factor structure. Larger samples with larger clusters are needed to conduct more robust, multilevel CFAs on the adapted PSAT. Third, the adapted PSAT for use in elementary schools was only validated in relation to one program, the scheduling of weekly classroom physical activity. Further research is needed to ensure it can be applied to assess sustainment of other public health programs in this setting. Fourth, there were slight differences in the implementation support strategies used across the two trials, which may introduce minor contextual differences that may impact on the measurement properties of the adapted PSAT. However, as the adapted PSAT was not measuring any aspect of the implementation strategies, such impacts should be minimal, if any. Finally, the adapted PSAT was designed to be completed by stakeholders with intimate knowledge of the organisational structures and factors, which may make it difficult for frontline staff to confidently and accurately complete all items in this measure [15]. Separate but complementary measures for administrative/leadership and frontline staff may be needed to ensure a comprehensive assessment of sustainment determinants is achieved.

5. Conclusions

If we are to ensure the long-term delivery and continued benefits of public health programs in schools, a clear understanding of the determinants of sustainment for school-based public health programs are needed, which is reliant on reliable, valid and pragmatic measures. Current attempts to adapt the PSAT for the elementary school setting among classroom teachers have been mixed, with evidence of internal reliability but mixed evidence of interrater reliability and validity. Future efforts to develop measures that are relevant and psychometrically robust for the elementary school setting and specific to the multiple end-user groups responsible for the delivery and governance of school-based public health programs are needed.

Supplementary Materials: The following are available online at www.mdpi.com/article/10.3390/ijerph18211414/s1, Table S1: Comparison of the original PSAT items to the adapted items for use in the elementary school setting.

Author Contributions: Conceptualization: A.H., L.W. and N.N.; data curation, A.S., C.L. and N.N.; formal analysis: A.H.; funding acquisition, L.W. and N.N.; investigation: A.H., A.S., R.C.S., C.L., L.W. and N.N.; methodology: A.H., R.C.S. and N.N.; project administration, A.S., C.L., L.W. and N.N.; resources, L.W. and N.N.; supervision, L.W. and N.N.; writing—original draft: A.H.; writing—review & editing: A.H., A.S., R.C.S., C.L., L.W. and N.N. All authors have read and agreed to the published version of the manuscript.

Funding: Authors N.N. and L.W. receive salary support from Hunter New England Local Health District, which contributes funding to the project outlined in this study. Data were collected from

two trials: one was funded by the National Health and Medical Research Council (NHMRC) Partnership Project grant (APP1133013). As part of the NHMRC Partnership Grant funding

Arrangement for this study, the following partner organisations also contribute funds: Hunter New England Local Health District and the NSW Health Office of Preventive Health. The second study from which data was used was supported by supported by Hunter Medical Research Institute (HMRI), Hunter Children's Research Foundation (HCRF), and Hunter New England Population Health. At the time of this study, N.N. was supported by a NHMRC TRIP Fellowship (APP1132450) and a Hunter New England Clinical Research Fellowship; L.W. was supported by a NHMRC Career Development Fellowship (APP1128348), Heart Foundation Future Leader Fellowship (101175) and a Hunter New England Clinical Research Fellowship. R.C.S. was supported by an American Cancer Society Research Scholar Grant (RSG-17-156-01-CPPB).

Institutional Review Board Statement: Human Research Ethics approval was obtained from Hunter New England (no. 06/07/26/4.04), the University of Newcastle (no. H-2008-0343) and relevant school bodies.

Informed Consent Statement: All participating schools provided informed written consent to take part, while return of a completed study survey by participating teachers was taken as implied consent to take part.

Data Availability Statement: Data and materials used in this study are available from the study team on reasonable request.

Acknowledgments: The authors would like to acknowledge members of the advisory group and evaluation team who contributed to the larger trials from which data for this study were obtained, including Patti-Jean Naylor, Angie Craddock, Karen Gillham, Benjamin Elton, Penny Reeves, Christopher Oldmeadow, Kirsty Hope, Bernadette Duggan, James Boyer, Matthew Pettett, Nicole McCarthy, John Wiggers, Adrian Bauman, Chris Rissell, Rachel Sutherland, Christophe Lecathelinais and Rebecca Jackson. We would also like to acknowledge the schools and teachers who took part in the study.

Conflicts of Interest: Authors N.N. and L.W. receive salary support from Hunter New England Local Health District, which contributes funding to the project outlined in this study. All other authors declare no conflicts of interest. Of the two trials from which data were obtained, one was funded by the National Health and Medical Research Council (NHMRC) Partnership Project grant (APP1133013) and the other was supported by supported by Hunter Medical Research Institute (HMRI), Hunter Children's Research Foundation (HCRF), and Hunter New England Population Health. NHMRC, HMRI and HCRF had no role in the design, data collection or analyses or writing of the manuscript. Individuals in positions that are funded by the partner organisations Hunter New England Local Health District (outlined above) had a role in the study design, data collection, analysis of data, interpretation of data and dissemination of findings.

References

1. Jacob, C.M.; Baird, J.; Barker, M.; Cooper, C.; Hamson, M. *The Importance of a Life Course Approach to Health: Chronic Disease Risk From Preconception through Adolescence and Adulthood: White Paper*; World Health Organization; 2017. Available online: <https://www.who.int/life-course/publications/life-course-approach-to-health.pdf> (accessed 18 March 2020).
2. World Health Organisation (WHO). *Global Status Report on Non-Communicable Diseases 2014*; World Health Organisation: Geneva, Switzerland, 2014.
3. World Health Organisation (WHO). *Global Action Plan for the Prevention and Control of Non-Communicable Diseases 2013–2020*; Geneva, Switzerland, 2013.
4. Langford, R.; Bonell, C.P.; Jones, H.E.; Poulou, T.; Murphy, S.M.; Waters, E.; Komro, K.A.; Gibbs, L.F.; Magnus, D.; Campbell, R. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. *Cochrane Database Syst. Rev.* **2014**, CD008958. <https://doi.org/10.1002/14651858.CD008958.pub2>.
5. Lima-Serrano, M.; Lima-Rodriguez, J.S. Impact of school-based health promotion interventions aimed at different behavioral domains: A systematic review. *Gac. Sanit.* **2014**, *28*, 411–417. <https://doi.org/10.1016/j.gaceta.2014.05.003>.
6. Dabravolskaj, J.; Montemurro, G.; Ekwaru, J.P.; Wu, X.Y.; Storey, K.; Campbell, S.; Veugelers, P.J.; Ohinmaa, A. Effectiveness of school-based health promotion interventions prioritized by stakeholders from health and education sectors: A systematic review and meta-analysis. *Prev. Med. Rep.* **2020**, *19*, 101138. <https://doi.org/10.1016/j.pmedr.2020.101138>.
7. Kriemler, S.; Meyer, U.; Martin, E.; van Sluijs, E.M.; Andersen, L.B.; Martin, B.W. Effect of school-based interventions on physical activity and fitness in children and adolescents: A review of reviews and systematic update. *Br. J. Sports Med.* **2011**, *45*, 923–930. <https://doi.org/10.1136/bjsports-2011-090186>.

8. Jacob, C.M.; Hardy-Johnson, P.L.; Inskip, H.M.; Morris, T.; Parsons, C.M.; Barrett, M.; Hanson, M.; Woods-Townsend, K.; Baird, J. A systematic review and meta-analysis of school-based interventions with health education to reduce body mass index in adolescents aged 10 to 19 years. *Int. J. Behav. Nutr. Phys. Act.* **2021**, *18*, 1. <https://doi.org/10.1186/s12966-020-01065-9>.
9. Neil-Sztramko, S.E.; Caldwell, H.; M., D. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. *Cochrane Database Syst. Rev.* **2021**, *9*. <https://doi.org/10.1002/14651858.CD007651.pub3>.
10. Messing, S.; Rutten, A.; Abu-Omar, K.; Ungerer-Rohrich, U.; Goodwin, L.; Burlacu, I.; et al. How can physical activity be promoted among children and adolescents? A systematic review of reviews across settings. *Front. Public Health* **2019**, *7*. <https://doi.org/10.3389/fpubh.2019.00055>.
11. Pate, R.R.; Trilk, J.L.; Wonwoo, B.J.W. Policies to increase physical activity in children and youth. *J. Exerc. Sci. Fit.* **2011**, *9*, 1–14.
12. Wiltsey Stirman, S.; Kimberly, J.; Cook, N.; Calloway, A.; Castro, F.; Charns, M. The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implement. Sci.* **2012**, *7*, 17. <https://doi.org/10.1186/1748-5908-7-17>.
13. Shelton, R.C.; Cooper, B.R.; Stirman, S.W. The Sustainability of Evidence-Based Interventions and Practices in Public Health and Health Care. *Annu. Rev. Public Health* **2018**, *39*, 55–76. <https://doi.org/10.1146/annurev-publhealth-040617-014731>.
14. Chambers, D.A.; Glasgow, R.; Stange, K. The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implement. Sci.* **2013**, *8*, 117.
15. Moullin, J.C.; Sklar, M.; Green, A.; Dickson, K.S.; Stadnick, N.A.; Reeder, K.; Aarons, G.A. Advancing the pragmatic measurement of sustainment: A narrative review of measures. *Implement. Sci. Commun.* **2020**, *1*, 76. <https://doi.org/10.1186/s43058-020-00068-8>.
16. Proctor, E.; Luke, D.; Calhoun, A.; McMillen, C.; Brownson, R.; McCrary, S.; Padek, M. Sustainability of evidence-based healthcare: Research agenda, methodological advances, and infrastructure support. *Implement. Sci.* **2015**, *10*, 88. <https://doi.org/10.1186/s13012-015-0274-5>.
17. Ament, S.M.; de Groot, J.J.; Maessen, J.M.; Dirksen, C.D.; van der Weijden, T.; Kleijnen, J. Sustainability of professionals' adherence to clinical practice guidelines in medical care: A systematic review. *BMJ Open* **2015**, *5*, e008073. <https://doi.org/10.1136/bmjopen-2015-008073>.
18. Herlitz, L.; MacIntyre, H.; Osborn, T.; Bonell, C. The sustainability of public health interventions in schools: A systematic review. *Implement. Sci.* **2020**, *15*, 4. <https://doi.org/10.1186/s13012-019-0961-8>.
19. Scheirer, M.A. Is Sustainability Possible? A Review and Commentary on Empirical Studies of Program Sustainability. *Am. J. Eval.* **2016**, *26*, 320–347. <https://doi.org/10.1177/1098214005278752>.
20. Cassar, S.; Salmon, J.; Timperio, A.; Naylor, P.J.; van Nassau, F.; Contardo Ayala, A.M.; Koorts, H. Adoption, implementation and sustainability of school-based physical activity and sedentary behaviour interventions in real-world settings: A systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2019**, *16*, 120. <https://doi.org/10.1186/s12966-019-0876-4>.
21. Shoosmith, A.; Hall, A.; Wolfenden, L.; Shelton, R.C.; Powell, B.J.; Brown, H.; McCrabb, S.; Sutherland, R.; Yoong, S.; Lane, C.; et al. Barriers and facilitators influencing the sustainment of health behaviour interventions in schools and childcare services: A systematic review. *Implement. Sci.* **2021**, *16*, 62. <https://doi.org/10.1186/s13012-021-01134-y>.
22. McLoughlin, G.M.; Allen, P.; Walsh-Bailey, C.; Brownson, R.C. A systematic review of school health policy measurement tools: Implementation determinants and outcomes. *Implement. Sci. Commun.* **2021**, *2*, 67. <https://doi.org/10.1186/s43058-021-00169-y>.
23. Lewis, C.C.; Fischer, S.; Weiner, B.J.; Stanick, C.; Kim, M.; Martinez, R.G. Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implement. Sci.* **2015**, *10*, 155. <https://doi.org/10.1186/s13012-015-0342-x>.
24. Mettert, K.; Lewis, C.; Dorsey, C.; Halko, H.; Weiner, B. Measuring implementation outcomes: An updated systematic review of measures' psychometric properties. *Implement. Res. Pract.* **2020**, *1*. <https://doi.org/10.1177/2633489520936644>.
25. McIntosh, K.M.L.D.; Hume, A.E.; Doolittle, J.; Vincent, C.G.; Horner, R.H.; Ervin, R.A. Development and initial validation of a measure to assess factors related to sustainability of school-wide positive behavior support. *J. Posit. Behav. Interv.* **2011**, *13*, 208–218. <https://doi.org/10.1177/1098300710385348>.
26. Luke, D.A.; Calhoun, A.; Robichaux, C.B.; Elliott, M.B.; Moreland-Russell, S. The Program Sustainability Assessment Tool: A new instrument for public health programs. *Prev. Chronic Dis.* **2014**, *11*, 130184. <https://doi.org/10.5888/pcd11.130184>.
27. Schell, S.F.; Luke, D.A.; Schooley, M.W.; Elliott, M.B.; Herbers, S.H.; Mueller, N.B.; et al. Public Health Program Capacity for Sustainability: A New Framework. *Implement. Sci.* **2013**, *8*, 15.
28. Washington University in St Louis. PSAT Program Sustainability Assessment Tool: Frequently Asked Questions. Available online: <https://sustaintool.org/psat/assess/> (accessed on 20 October 2021).
29. Boateng, G.O.; Neilands, T.B.; Frongillo, E.A.; Melgar-Quinonez, H.R.; Young, S.L. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front. Public Health* **2018**, *6*, 149. <https://doi.org/10.3389/fpubh.2018.00149>.
30. Terwee, C.B.; Bot, S.D.; de Boer, M.R.; van der Windt, D.A.; Knol, D.L.; Dekker, J.; Bouter, L.M.; de Vet, H.C. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **2007**, *60*, 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
31. Nathan, N.K.; Sutherland, R.L.; Hope, K.; McCarthy, N.J.; Pettett, M.; Elton, B.; Jackson, R.; Trost, S.G.; Lecathelinais, C.; Reilly, K.; et al. Implementation of a School Physical Activity Policy Improves Student Physical Activity Levels: Outcomes of a Cluster-Randomized Controlled Trial. *J. Phys. Act. Health* **2020**, *17*, 1009–1018. <https://doi.org/10.1123/jpah.2019-0595>.

32. Nathan, N.; Hall, A.; McCarthy, N.; Sutherland, R.; Wiggers, J.; Bauman, A.E.; Rissel, C.; Naylor, P.J.; Cradock, A.; Lane, C.; et al. Multi-strategy intervention increases school implementation and maintenance of a mandatory physical activity policy: Outcomes of a cluster randomised controlled trial. *Br. J. Sports Med.* **2021**. <https://doi.org/10.1136/bjsports-2020-103764>.
33. Wiggers, J.; Wolfenden, L.; Campbell, E.; Gillham, K.; Bell, C.; Sutherland, R.; Hardy, L.; King, L.; Grunseit, A.; Milat, A.; et al. *Good for Kids, Good for Life 2006–2010 Evaluation Report*; NSW Ministry of Health: Sydney, Australia, 2013.
34. NSW Government. *Rationale for Change; Sport and Physical Activity Policy—Revised 2015*; NSW Department of Education: Bankstown, Australia; 2015.
35. Mokkink, L.B.; Prinsen, C.A.C.; Patrick, D.L.; Alonso, J.; Bouter, L.M.; de Vet, H.C.W.; et al. *COSMIN Methodology for Systematic Reviews of Patient-Reported Outcome Measures (PROMs): User Manual*; The Netherlands, 2018. Available online: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf (accessed online: 9 Nov 2020)
36. O'Rourke, N.; Hatcher, L. *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling: Second Edition*; SAS Institute Inc: Cary, NC, USA, 2013.
37. Fabrigar, L.R.; Wegener, D.T.; MacCallum, R.C.; Strahan, E. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychol. Methods* **1999**, *4*. <https://doi.org/10.1037/1082-989X.4.3.272>.
38. Hox, J. *Multilevel Analysis: Techniques and Applications*; Lawrence Erlbaum Associates Inc.: Mahwah, NJ, USA, 2002.
39. Huang, F. Conducting Multilevel Confirmatory Factor Analysis Using R. Available online: <http://faculty.missouri.edu/huangf/data/mcfa/MCFA%20in%20R%20HUANG.pdf> (accessed on 16 Mar 2021).
40. Hooper, D.; Coughlan, J.; MR., M. Structural Equation Modelling: Guidelines for Determining Model Fit. *Electron. J. Bus. Res. Methods* **2008**, *6*, 53–60. <https://doi.org/10.21427/D7CF7R>.
41. King, D.K.; Gonzalez, S.J.; Hartje, J.A.; Hanson, B.L.; Edney, C.; Snell, H.; Zoorob, R.J.; Roget, N.A. Examining the sustainability potential of a multisite pilot to integrate alcohol screening and brief intervention within three primary care systems. *Transl. Behav. Med.* **2018**, *8*, 776–784. <https://doi.org/10.1093/tbm/ibx020>.
42. Bliese, P. Within-group Agreement, Non-Independence, and Reliability. In *Multilevel Theory, Research, and Methods in Organizations*; Klein, K., Kozlowski, S., Eds. Jossey-Bass Inc.: San Francisco, CA, USA, 2000.
43. Fernandez, M.; Walker, T.; Weiner, B.; Calo, W.; Liang, S.; Risendal, B.; Friendman, D.; Tu, S.; Williams, R.; Jacobs, S.; et al. Developing measures to assess constructs from the Inner Setting domain of the Consolidated Framework for Implementation Research. *Implement. Sci.* **2018**, *13*, 52.
44. Cawaling, E.J.B.; Cunanan, D.U.; Bernarte, R.P. Sustainability Capacity of HIV Programs in National Capital Region, Philippines. *Int. J. Public Health Sci. (IJPHS)* **2018**, *7*. <https://doi.org/10.11591/ijphs.v7i2.12408>.
45. Llauro, E.; Aceves-Martins, M.; Tarro, L.; Papell-Garcia, I.; Puiggros, F.; Prades-Tena, J.; Kettner, H.; Arola, L.; Giralt, M.; Sola, R. The "Som la Pera" intervention: Sustainability capacity evaluation of a peer-led social-marketing intervention to encourage healthy lifestyles among adolescents. *Transl. Behav. Med.* **2018**, *8*, 739–744. <https://doi.org/10.1093/tbm/ibx065>.
46. Cabassa, L.J.; Stefancic, A.; Bochicchio, L.; Tuda, D.; Weatherly, C.; Lengnick-Hall, R. Organization leaders' decisions to sustain a peer-led healthy lifestyle intervention for people with serious mental illness in supportive housing. *Transl. Behav. Med.* **2021**, *11*, 1151–1159. <https://doi.org/10.1093/tbm/ibaa089>.
47. Nazar, H.; Nazar, Z. Community pharmacy minor ailment services: Pharmacy stakeholder perspectives on the factors affecting sustainability. *Res. Soc. Adm. Pharm.* **2019**, *15*, 292–302. <https://doi.org/10.1016/j.sapharm.2018.04.036>.
48. Stoll, S.; Janevic, M.; Lara, M.; Ramos-Valencia, G.; Stephens, T.B.; Persky, V.; Uyeda, K.; Ohadike, Y.; Malveaux, F. A Mixed-Method Application of the Program Sustainability Assessment Tool to Evaluate the Sustainability of 4 Pediatric Asthma Care Coordination Programs. *Prev. Chronic. Dis.* **2015**, *12*, E214. <https://doi.org/10.5888/pcd12.150133>.
49. Calhoun, A.; Mainor, A.; Moreland-Russell, S.; Maier, R.; Brossart, L.; Luke, D. Using the Program Sustainability Assessment Tool to assess and plan for sustainability. *Prev. Chronic. Dis.* **2014**, *11*. <https://doi.org/10.5888/pcd11.130185>.