Helping people access and use quality evidence

NEW ZEALAND HEALTHY SCHOOL LUNCH PILOT

KA ORA, KA AKO INTERIM EVALUATION

Evaluation Appendices and Technical Report

17 May 2021

STANDARD OF PROOF

CONTENTS

Appendix A: Methodology

The overall methodology was designed to estimate the (net) effects of the pilot on the expected outcomes, with interim results required at the end of the first year and a final report in early 2021. The pilot was expanded during this same period, and although this expansion was out of scope for the current work, the Ministry of Education plans to evaluate this expanded programme. The Ka Ora, Ka Ako Pilot Evaluation therefore should be considered alongside any future evaluations of the programme.

Learner food diaries

Learners were asked to complete food diaries over a five-day period in September, and again in November. The details of this learner food diary are provided here.

Instrument

Apart from collating demographic information about the children (age, sex and ethnicity), learner food diaries were developed to measure change in learners' wellbeing, satiety, food availability and consumption. The learner food diary consisted of different measures. These were:

1. Teddy-the-Bear satiety scale[1]: The Teddy-the-Bear scale is a short (one question), easy to complete measure of satiety appropriate for young children aged 5 and above. It has been found to be a valid instrument measuring satiety, and found to indicate no significant variance among boys and girls at different age ranges among pre-school children in Great Britian. There are currently no normative data available in New Zealand, and differential item functioning needs to be determined within the context.
2. WHO-5 (World Health Organisation 5-Wellbeing Index)[2]: The WHO-5 is an internationally recognised measure of mental wellbeing, and among the most widely used questionnaires assessing subjective psychological wellbeing in the world. The wellbeing index is a short and simple assessment, and is appropriate for children aged 9 and older. It is based on five individual items, measuring positive mood (good spirits, relaxation), vitality (being active, waking up fresh and rested), and general interest (being interested in things). The WHO-5 has been found to have high clinimetric validity, can be used as an outcome measure balancing the wanted and unwanted effects of treatments, is a sensitive and specific screening tool for depression and its applicability across study fields is very high[3].

[1] Bennett, C., & Blissett, J. (2014). Measuring hunger and satiety in primary school children. Validation of a new picture rating scale. Appetite, 78, 40–48.
[2] https://www.psykiatri-regionh.dk/who-5/Pages/default.aspx
[3] Christian Winther Topp, Søren Dinesen Østergaard, Susan Søndergaard, Per Bech (2015), *The WHO-5 Well-Being Index: a systematic review of the literature,* Psychotherapy and Psychosomatics, 84(3):167-76.

3. PedsQL™ (Pediatric Quality of Life Inventory, Generic Core scale, Version 4.0 – Australian English, Child self report 8-12[4]): PedsQL™ is a widely used and brief measure of quality of life. It is a multidimensional scale measuring physical, emotional, social and school functioning of the child. It estimates the quality of life as a total score as well as physical health and psychosocial health. The PedsQL™ has been found to be a valid and reliable scale, applicable in clinical trials, research, clinical practice, school health settings, and community populations.[5] There are normative New Zealand data available, but for specific populations.

4. Food availability and consumption: After reviewing and testing different instruments with children, it was determined necessary to create a measure that could be used with children aged 5 to 12 and could estimate the availability and consumption of different food types aligning with the Ministry of Health nutrition guidelines. The food items focused on vegetables and fresh fruit, or items that were consistently labelled as "green" or "healthy" in the guidelines irrespective of how they were prepared; as well as items that were consistently labelled as "unhealthy" or "red" options such as snacks and sweet items. The specific food item choices were considered easy to identify, even for children, and were originally derived from the food choices available in the New Zealand Child Nutrition survey (Food Frequency Questionnaire), and supplimented in discussion with the Ministry of Health.

The food diaries were designed with young learners in mind, and Māori translations were provided alongside the English text. Pictures were used wherever relevant given the age of the children. The learner food diaries were provided as booklets to classroom teachers.

In the leaner food diaries, demographic information, satiety and wellbeing was collected on Day 1, whereas satiety, food availability and consumption were collected on Days 2, 3, 4 and 5 (c.f. Appendix C). The diaries therefore provide measures over a five-day period for satiety, a four-day period for food availability and consumption, while learners are asked to reflect on the previous 2 weeks (WHO-5) or month (PEDSQL™) for the wellbeing measures. These measures were captured in both the September and November learner diaries.

Two rounds of user testing were adopted, in addition to incorporating feedback and views from the Ministry of Health. Different user-testing techniques were used, including researchers observing classrooms of respondents as well as discussing the instrument in small focus groups of the target audiences in both Māori and English. Modifications to the instrument were made, as well as identifying the most appropriate data collection procedure (i.e. class facilitation).

[4] The license to use the instruction was purchased specifically for this evaluation.
[5] J W Varni 1, M Seid, P S Kurtin (2001). PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations, Medical Care 2001, Aug: 39(8): 800-12.

The diaries were facilitated by the classroom teachers, which was considered the most reasonable approach given COVID-19. Introducing external researchers to facilitate the classroom exercise wasn't appropriate given the likely ongoing health concerns and anxieties felt by schools, their whānau and communities. Procedures were developed for classroom teachers, and multiple online opportunities were provided for them to ask questions about facilitating the learner food diaries. The protocols included:

1. Teachers provide each learner with a food diary, and learners put their name on their diaries.
2. Teachers facilitate the diaries with the class immediately following lunch over 5 subsequent days and within 60 minutes of beginning lunch, helping those learners who may need extra support.
3. Learners remove their diary entries from their diary daily, and teachers collect the dairy entries and seal them in provided envelopes.
4. Teachers confirm the protocols have been followed on the envelope, and provide learner enrolment in the class and attendance for the day.
5. Teachers collect the diaries and redistribute these to the named learners the next day.

Learner food diaries were printed and couriered to each participating class in August. Administration instructions were also provided to each classroom teacher with their bundle of diaries. Sealable envelopes for each of the five days were also provided to the teacher. On each envelope were instructions and questions about class attendance, as well as questions confirming the administration procedures were followed.

Approach

A longitudinal, Before-After Control-Impact (BACI) design was adopted for this interim evaluation of the pilot, whereas the "BA" references the longitudinal nature of the design and the "CI" references the comparison groups.

A longitudinal approach provided evidence about the programme's effect on specific learners, such as Māori and those most in need. Whereas a cross-sectional approach would determine if Ka Ora, Ka Ako made systematic improvements to food security, attendance and wellbeing at the group level rather than for individuals. Although the cost and burden is higher in longitudinal designs, it is expected that the Ka Ora, Ka Ako programme will impact more on those with greatest need, and the provision for all learner lunches will reduce the stigma associated with poverty and needing food. Therefore, a longitudinal design was considered the most appropriate given the context of the programme, while also providing a more efficient estimate of mean change when compared with repeated cross-sectional samples"[6].

[66] Hayes and Moulton, 2017. Cluster randomised trials, pg. 182.

The learner food diaries included a cover page which allowed learners to name their specific books. However, this information was not provided to the evaluation team to ensure the anonymity of the learners was protected. To track change relevant to each person over time, each page of the diary was stamped with a number unique to the book. Each page of the diary was perforated so the learner could remove the anonymous diary entry for the day and submit that into a sealed envelope managed by the teacher. The unique numbers allowed the evaluation team to track progress among each learner according to their diary entries.

Five days of data were collected in September, before the programme began, and another five days of data were collected in November, after the programme began for many schools. Although this is a short period of time, it was determined as appropriate given the reporting needs. The expectation was that improvements in food availability and satiety should be evident by this time whereas wellbeing may take longer to achieve.

Similar comparison groups were also necessary in this interim evaluation, given simple pre- and post-test comparisons (without a counterfactual) would likely overestimate the programme effects.[7] This may be due to the seasonal affects relevant to food availability, here when comparing the availability and consumption of fresh fruit and vegetables between the spring (baseline assessment) and summer seasons (post-assessment). As such, comparison groups or at least time series data are required for a robust impact evaluation.

We compared results from those schools providing Ka Ora, Ka Ako lunches with those not (yet) providing lunches to their learners. More specifically, we collected information from all schools at fixed points in time (e.g. September and November), making a 'vertical' slice of the experiences at specific times, and compared the results between those schools delivering lunches before the follow up data collection (November) with those schools delivering lunch in 2021.

This stepped-wise design was intended to capture similar numbers of schools in both the control and impact groups. However, as the programme rolled out, the majority of the schools were able to start earlier, meaning our control group is smaller than the treatment group.

What is even more important is that the schools starting earlier (treatment schools) are not systematically different from those starting later (control group schools). This will be assessed as part of the post assessment.

Sampling strategy

The sampling strategy focussed on gaining a representative view of the primary and intermediate learners in the Hawkes Bay and Bay of Plenty regions involved in the Ka Ora, Ka Ako pilot schools. Cluster sampling is often used when it's not practical to randomly sample from a population, such as here. Although it's not practical to sample

[7] Aiken et al., 1998, as cited in Rossi et al. 2004: 298).

across all learners in the Hawke's Bay and Bay of Plenty regions, it is possible to sample clusters of these learners within schools and classes. Albeit more complex, we adopted the more practical one-stage cluster sampling approach focused on groups of classes.

There were 129 schools providing for primary and intermediate aged learners in the Hawke's Bay and Bay of Plenty regions deemed eligible for the programme. Eligibility was determined by the programme according to location (region), level of deprivation within the school as determined by the equity index, and information from Ministry of Education regional offices. This sample frame was complete, including only those primary and intermediate schools in these two regions that had been offered and accepted a Ka Ora, Ka Ako place, and all 129 schools were asked if they would be willing to contribute to an evaluation in different ways.

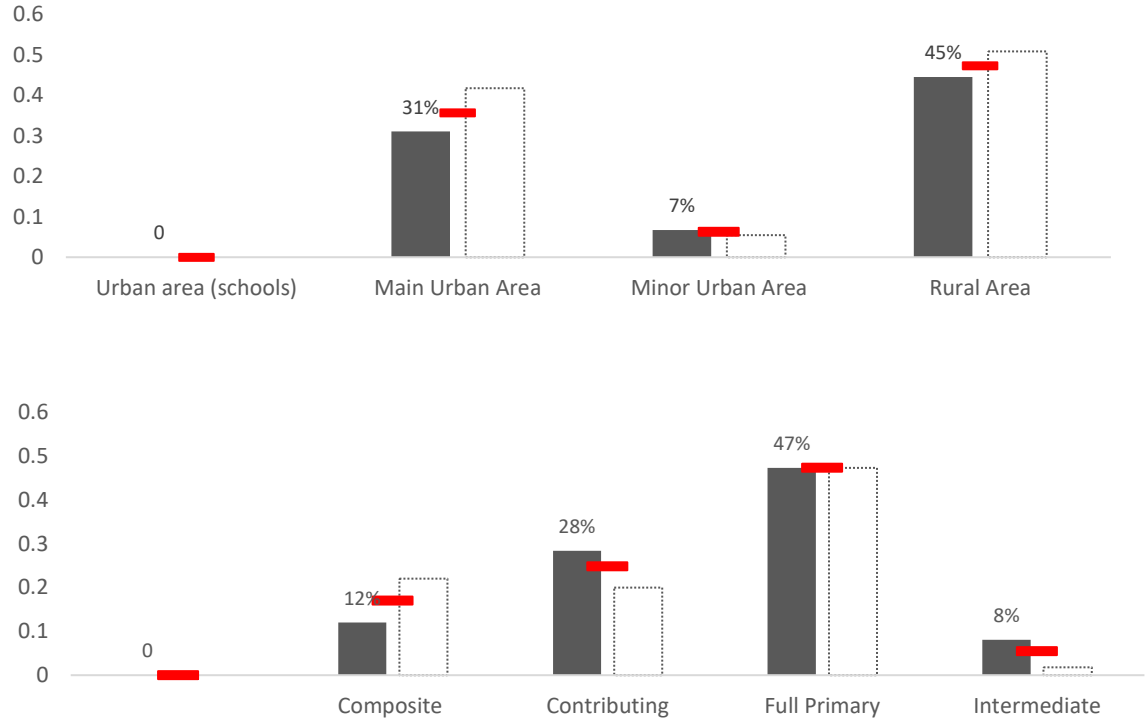
In total 74 of the eligible 129 schools responded to the initial invitation, which helped determine the approach for data capture. The responses showed that 7% (n=5) were unwilling to participate in the evaluation, while the other schools suggested that:

1. School teachers would be willing to facilitate the food diaries with learners (1%, n=1)
2. School teachers would be willing to facilitate the food diaries with learners, but they did not want an external evaluation team to come into the school (15%, n=11);
3. They did not want their teachers to facilitate the food diaries with learners, but they would welcome an external team coming into the school (9%, n=7); or
4. They would either be willing to have the evaluation team visit or the teachers facilite the diaries (68%, n=50).

Given the preferences, along with the likely occurrence of COVID-19 levels impacting on the permission for an external visitor to visit the schools twice over two terms, it was decided to facilitate the diaries through school teachers. This meant that the evaluation was able to select from the 62 schools (#1, #2, and #4 above) who would be open to this approach. The implications were that only willing classroom teachers could be engaged, rather than surveying the full school.

These 62 schools were mostly similar to the 129 eligible schools insofar as proportional makeup of rurality, region and organisation type. As shown below, there were no meaningful differences between those willing to take part in this way in comparison to those eligible schools when considering region ($X^2_3 = 2.049$, $p > 0.05$), school type ($X^2_4 = 8.017$, $p > 0.05$) or school size ($F(1,71) = 0.239$, $p = 0.626$). However, there was a small difference in type of area ($X^2_3 = 8.697$, $p = 0.034$), with a marginally greater representation in the secondary urban area than the eligible population; this was due to all of these schools responding to the initial request for information. Upon consideration, it was determined that using schools that were willing to participate and facilitate diaries, as indicated in the initial invitation to participate, would have a negligible impact on the sample demographics available for the evaluation.

Figure 1. Proportional makeup of schools willing (grey bar) vs. not willing (white bar) to take part in the evaluation in comparison to the eligible population (red line).



As such, the population sample frame included up to 62 primary and intermediate schools across the Hawke's Bay and Bay of Plenty regions that had indicated they would be open to participating in the evaluation by facilitating the diaries. However, a whole-of-school approach was not possible as schools nominated specific teachers who would be willing to contribute. When examining the coverage within these 62 schools, most schools (60%, n=37) nominated all of their classes whereas fewer schools (40%, n=25) were more selective about which classes would be willing to participate in the food diaries, nominating less than half of their learner population. Knowing satiety and wellbeing may be unique for different age groups, our sampling strategy focused on a single-stage cluster sampling approach, with stratification occurring at the year level. What this means is that school clusters would be selected to ensure similar representation across the age groups.

We estimated a medium effect size was achievable, and therefore required responses from approximately 3,000 learners across the eight year levels.[8] Year levels were grouped to accommodate classes providing for multiple Year levels using the estimates

[8] 312 randomly selected matched subject responses (Repeated Measures Analysis of Variance) and includes two measures (pre and post) and four between-subject factors totalling 24 levels (1-β =0.80; α =0.05, partial η2 = 0.02). However, in this cluster sampling approach, it is further necessary to account for the correlations between the outcomes and the clustering effects that are expected at the class level. The estimated design effect of satiety on age shown from the Term 1 assessment was 6.7 (ICC = 0.07), which is reasonably high. The sample size was adjusted to account for the cluster design effect expected within the age groups, we would require 2,100 participants. We estimated losing a portion of the sample due to absentisms or simple non-response at the learner, class and/or school level, and therefore sampled for 3,500 learners.

provided by schools[9]. This resulted in a sample target of approximately 1,500 learners for Years 1-4, and approximately 750 across Years 5-6 and 7-8. These were then grouped according to each stratum (Years 1-4, Years 5-6, Years 7-8), and randomised with these strata. The resultant list of classes and schools that would be invited to take part in the evaluation included 152 classes across 45 schools.

No sampling would occur within these 152 classes (i.e. a census), which meant we estimated reaching approximately 3,200 learners.

Data entry and cleaning

The quality of data is assured through coder-training and double data entry preformed on an adequate proportion of the data. A team of five individuals were trained as a group, and the coding rules were explained and discussed. The coding rules were available on a whiteboard, and data anomalies were identified as coding was underway, and discussed with the team to determine alignment with the established coding rules. Any anomilies were then exemplified on the whiteboard against the coding rules.

After data was entered into an excel spreadsheet. Double data entry is performed wherein the data is entered by two coders separately. The entry made by the second person helps in verification and reconciliation by identifying the transcription errors and discrepancies. This helps to get a cleaner database and ensures better consistency and a lower error rate. Manual double entry of data is defined as the definitive gold standard, and has been shown to exceed accuracy beyond visual inspections and read-aloud methods[10] [11].

The quality assurance process was undertaken while the 'first-pass' data entry was underway. Researchers randomly selected classes, being sure that they had not entered the specific class of data in the first pass, and then entered the data a second time. Any errors are automatically flagged and discussed as a team, and the errors and corrections documented. In total, 19.8% of My First Day and 10.7% of Days 2 to 5 class entries underwent a second pass data entry, resulting in 54 and 139 anomolies identified in My First Day and Days 2-5, respectively, between the two coders; this means between three to five anomalies were identified (and then corrected) for every 1,000 item responses entered. With the anomalies corrected to 19.8% of My First Day and 10.7% to Days 2 to 5, we are confident that 99.73% (±0.07%) of the My First Day and 99.53% (±0.11%) of Days 2 to 5 accurately reflect the paper-based data submitted by the learner.

The process for entering the post-assessment was similar, including initial training and parallel double manual entry. The quality assurance results were in summary, 15.88% of

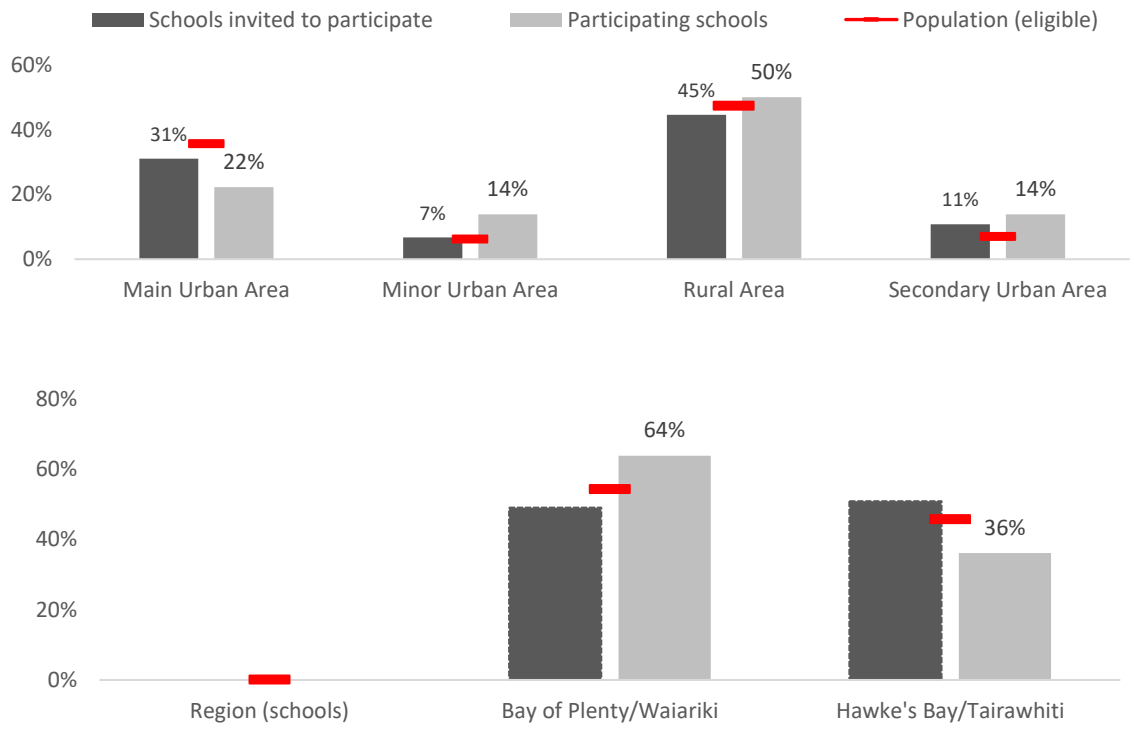
[9] The numbers of learners in a classroom was estimated in some cases where this information was not provided by the school.
[10] Ohmann C, Kuchinke W, Canham S, Lauritsen J, Salas N, et al. (2011). Standard requirements for GCP-compliant data management in multinational clinical trials. Trials 12: 85
[11] Kimberly A BarchardYevgeniya Verenikina (2013). Improving data accuracy: Selecting the best data checking technique. Computers in Human Behavior 29(5):1917–1922.

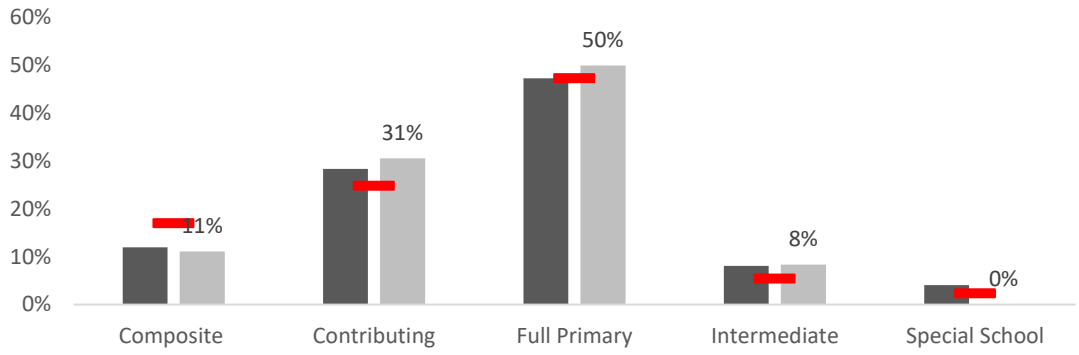
My First Day and 11.45% of Days 2 to 5 class entries underwent a second pass data entry, resulting in 14 and 78 anomolies identified in My First Day and Days 2-5, respectively, between the two coders; this means between one to two anomalies were identified (and then corrected) for every 1,000 item responses entered; an overall improvement in data entry accuracy rates. With the anomalies corrected to 15.88% of My First Day and 11.45% to Days 2 to 5, we are confident that 99.88% (±0.07%) of the My First Day and 99.80% (±0.11%) of Days 2 to 5 accurately reflect the paper-based data submitted by the learners in week 2.

Sample size and response rate

At baseline, 38 schools returned baseline data across 134 classes, which represents a response rate of 84% at the school level and 88% at the class level. Competing priorities within the school was the most common noted reason for dropping out of the evaluation. As shown below, there were no meaningful differences between those that took part in the evaluation (n=36) and those that dropped out (n=9) when considering region ($X^2_1 = .024$, $p > 0.05$), school type ($X^2_3 = 2.771$, $p > 0.05$), and type of area ($X^2_3 = 1.667$, $p > 0.05$), but it appeared that larger schools were able to participate more easily than smaller schools ($F(1,44) = 6.148$, $p < 0.017$). We did not weight by school size as wanted the data to represent the learner rather than the school.

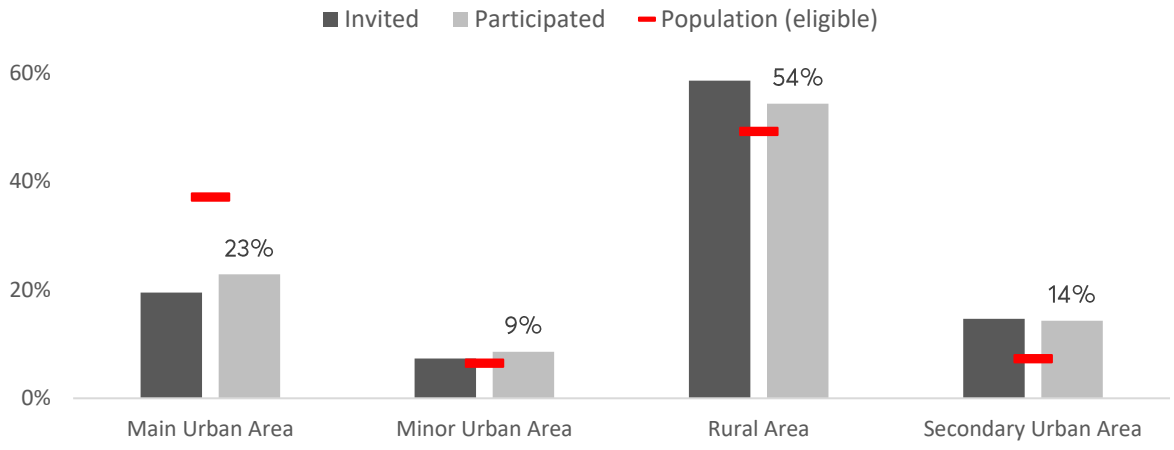
Figure 2: Baseline sample representation, comparing schools that participated (grey bar) with those that dropped out of the interim evaluation (dark grey bars) in relation to the overall population

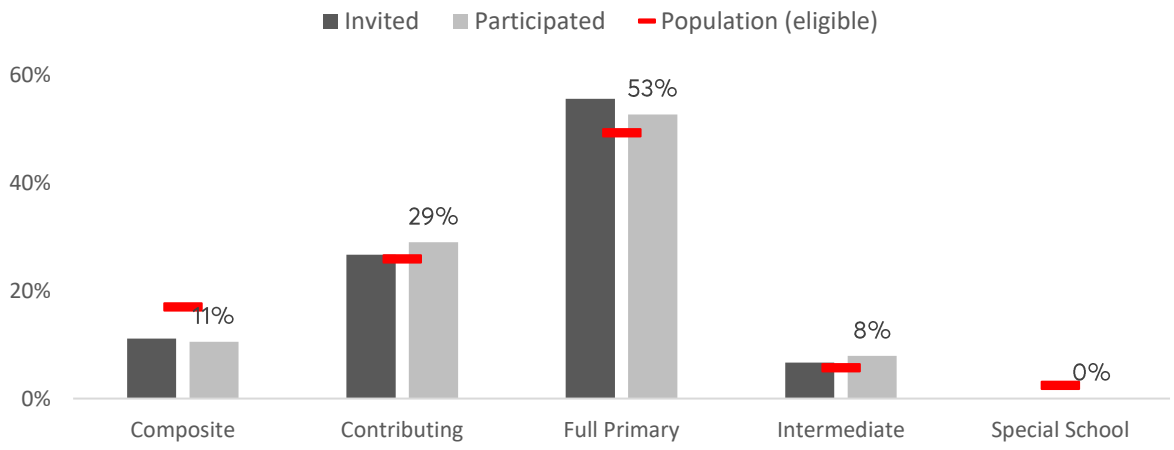




At the class level, the response rate was notably high. Class enrolment sizes were confirmed by the teacher returning the envelopes. The information allowed us to calculate a response rate for each day across all classes, showing that an average rate of response across the classes and the five days was 85% of class enrolments. As shown below, this response rate trended marginally downwards over the five days which was likely due to learners not being present over the full five-day period.

Figure 3: Final data sample representation, comparing schools that participated (grey bar) with those that dropped out of the interim evaluation (dark grey bars) in relation to the overall population.



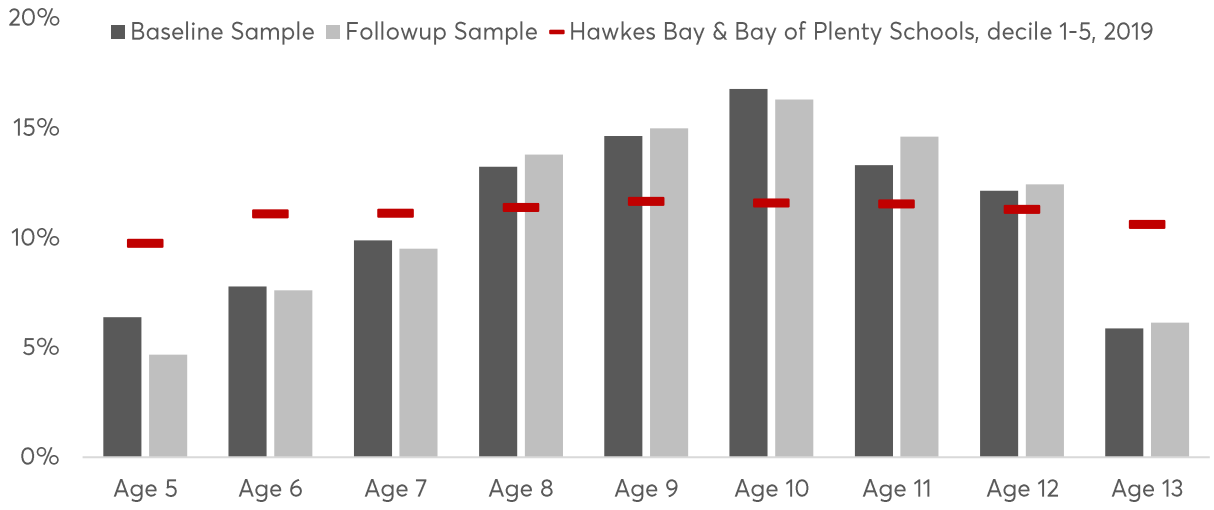


From the eligible population of 129 decile 1-5 schools in the Hawkes Bay and Bay of Plenty region, 65% were eliminated due to sampling and then a further 10% were lost due to respondent dropout. A smililar pattern of attrition was observed for students where the eligible population of 20,798 was reduced by 84% due to sampling and then a futher 15% and 8% due to respondent completion and data cleaning respectively, resulting in a final sample of 2166 learners for the longitudional analysis.

Table 1: Rates of attrition across sample, response and cleaning stages

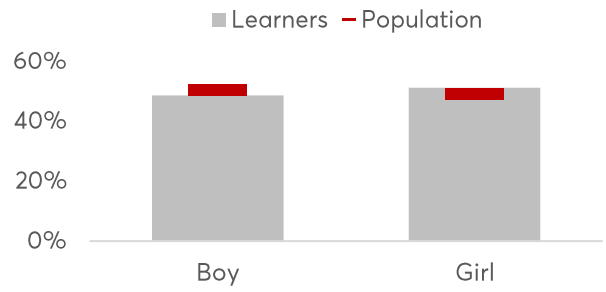
Attrition due to:	Sampling			Respondent completion		Data cleaning
	Eligible	Responded	Invited	Baseline	Followup	Clean
School	129	74	45	38	34	34
Class		217	153	134	116	111
Student	20,798[12]	11,118[13]	3290[14]	2775	2358	2166

Figure 4: Sample representation of age at baseline and followup in relation to Hawkes Bay & Bay of Plenty Schools, decile 1-5, 2019.



[12] School roll statistics for eligible schools
[13] Estimate based on school roll sizes
[14] Estimated based on school provided information and imputations (where class sizes were not provided by the school)

Figure 5: Sample representation of sex in relation to Hawkes Bay & Bay of Plenty Schools, decile 1-5, 2019.



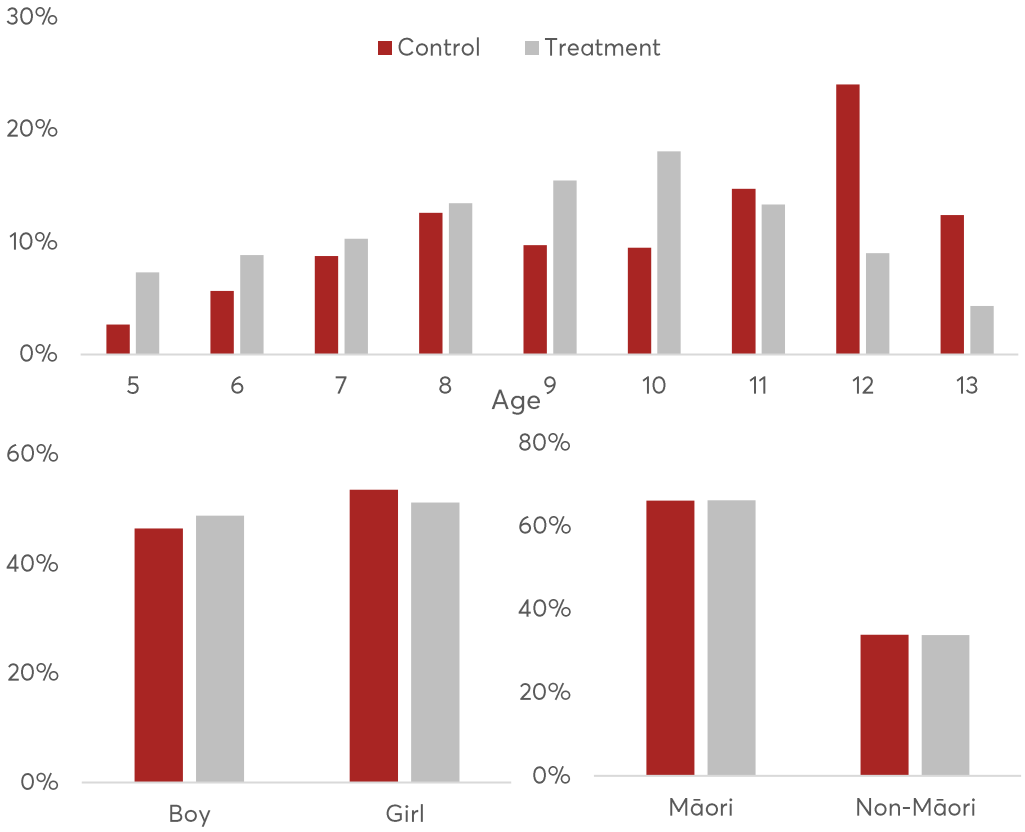
Sample subset comparisons (control vs. treatment; most disadvantaged groups)

When comparing both treatment and control groups, demographic variables sex and Māori had equal representation in both treatment and control schools. The distribution of age, however, differered between treatment and control schools. As such age was included as a potential confounder for the longitudinal modelling work.

- The interim evaluation also focussed on examining benefits for the full population of learners, which were expected to be positive but minimal. Given the smaller expected effect size and the pilot's whole-of-school approach, the sampling approach focussed on generalisability for the population of primary and intermediate learners for the pilot.

The demographic makeup of these groups are presented below, for ethnicity, age and sex.

Figure 6: Sample representation of control and treatment groups, by sex, Māori, and age.



It was further expected that those most disadvantaged learners (those 'most in need') would realise even larger benefits than this overall population, and as such, require fewer learners' data to detect statistical significance. These two groups of learners were those learners with:

- **Insufficient quantities of food** - defined here as the learners that were hungry after lunch for at least 50% of the days reported before the pilot; and
- **Fewer healthy options** - defined here as the learners that had snacks and sweets make up 60% or more of their *available*[15] lunch items while vegetables and fresh fruit items make up less than 60% of available lunch items.

Item analysis and imputation

The overall response rates for the closed response questions within the learner food diaries is provided below. As shown, sex and age were the items that attracted the highest response rates with nearly all respondents answering these questions. More than three-quarters of the eligible learners responded to the items in relation to the WHO-5 and the PEDS-QL.

[15] Availability rather than consumption is being used to identify these children as we wanted to omit the potential influence of 'picky' eaters and select only those that did not have this food available.

Table 2: Response rate for each item recorded in the Learner Food Diarys.

	Item	September	November
Day 1	demo_x1_sex	99.4%	94.9%
	demo_x2_maori	97.5%	
	demo_x3_pasifik	95.1%	
	demo_x4_pakeh	94.7%	
	demo_x5_age	98.9%	93.3%
All Days	sat_x6_teddy	96.4%	95.9%
Day 2-5	sat_x3_eatlunch	92.0%	91.8%
Ages 9+	who_x1_cheer	98.5%	97.4%
	who_x2_calm	98.5%	97.8%
	who_x3_active	98.5%	97.3%
	who_x4_fresh	98.3%	97.4%
	who_x5_daylife	98.2%	97.3%
Ages 8-12	peds_physical_x1_walk	95.9%	93.2%
	peds_physical_x2_run	96.7%	94.0%
	peds_physical_x3_sport	95.9%	93.0%
	peds_physical_x4_lift	96.3%	92.7%
	peds_physical_x5_bath	96.0%	93.4%
	peds_physical_x6_help	95.6%	92.3%
	peds_physical_x7_aches	95.1%	92.6%
	peds_physical_x8_lowen	96.4%	93.5%
	peds_emotional_x1_afraid	96.4%	93.8%
	peds_emotional_x2_sad	95.9%	93.2%
	peds_emotional_x3_angry	95.8%	92.3%
	peds_emotional_x4_sleep	96.1%	93.1%
	peds_emotional_x5_worry	96.1%	93.3%
	peds_social_x1_getalong	96.8%	93.5%
	peds_social_x2_friend	95.8%	92.6%
	peds_social_x3_tease	95.6%	92.7%
	peds_social_x4_do	95.7%	92.7%
	peds_social_x5_keepup	95.6%	93.4%
	peds_school_x1_payattention	95.8%	92.8%
	peds_school_x2_forget	95.6%	92.7%
	peds_school_x3_schoolwork	95.1%	92.2%
	peds_school_x4_sick	95.0%	92.5%
	peds_school_x5_doct	95.5%	92.8%

Imputation was completed only for age, where the mean age of fellow classmates in the class (who did record their age) was used for those learners with missing age.

The scoring procedures and completion rates are provided below for each assessment:

- Satiety Score: Responses on a 5-point satiety scale were transformed onto a scale ranging from 0 to 100. Average satiety scores were obtained as a mean across the recorded days of satiety.
- PEDSQL: Responses were transformed onto a scale ranging from 0 to 100, with higher scores indicating better quality of life. These scores were calculated for each construct (physical, emotional, social, school) when at least 50% of items were answered within the scale. The overall Quality of Life was estimated as the mean across these constructs only if all constructs were calculated.
- WHO-5: The overall wellbeing score was calculated as the sum of scores across the five questions, multiplied by four. The scores were calculated when all of the items were answered within the scale.

Response rates were not possible for the food item assessments, as the choice for learners was to indicate if each food item was consumed or available. This means that non-ciricled/crossed items may demonstrate either that the food item was not available or an item non-response. Although not ideal in terms of measurement, the approach was determined as appropriate given the age of the youngest respondents (5 years old) and context (Covid-19 engagement restrictions limiting our data capture approach). It is for this reason that the results are presented according to what was circled (consumed) and crossed (available but not consumed) but not presented according to items that were not available to the learner.

Error and robust estimates

The approach used a complex sampling design, which involved clustering at the school and class level. It is therefore not appropriate to calculate the standard errors using formulas that assume simple random sampling. Sampling errors were calculated incorporating the design effect (deff), which is the estimated gain or loss using the given sampling approach when compared to a simple random sample. Deff was estimated as 1+(m-1)*ICC, whereby m is the average number of learners in a class and ICC represents the intracluster correlation coefficient or the strength of the relationship within classes. The sampling error was calculated as 1.96 * sqrt (deff * p(1-p)/n).

The ICC when considering classes as the cluster was 0.118, and the design effect was 3.14 (m=19.2). This deff which was lower than expected when estimating our sample size requirements, meaning our samples should be sufficient to accommodate the requisite comparisons and statistical techniques. For future reference, sampling would have been equally as efficient if done at the school level given these data indicated ICC for schools was 0.0307, with a design effect of 3.02 (m=66.8); a census approach within the school will likely be a viable option when administering the diaries this way.

Confidence intervals (CIs) were computed using robust standard errors to accommodate the clustering effects expected to occur at the class level. These cluster robust estimates are reported throughout the report, wherever CIs are reported.

Sample weighting of the results was necessary so that the distribution of learners ages in the sample was representative of the overall population. Weighting was conducted

via post-stratification for the cross-sectional and prevalence statistics only to provide an accurate estimate of need. Weights were based[16] off Ministry data for learners aged 5 – 13 in Hawkes Bay and Bay of Plenty decile 1-5 schools, 2019 (the population), and learners in the followup food diaries (the sample). The learners ages, and their corresponding weights are given in the table below.

Analysis, modelling, and error estimation were done using R software and relevant packages[17].

Table 3: Weights by age

Age	5	6	7	8	9	10	11	12	13
Weights	2.089	1.458	1.170	0.825	0.778	0.712	0.790	0.908	1.729

Net Impact estimate (learners in control vs. treatment schools)

For the longitudinal analysis the net-effect of the pilot was estimated using linear mixed modelling. Here the cluster design was accounted for by including a classroom level random effect. The treatment indicator variable and confounding factor were modelled as fixed effects. The model parameters were estimated using the restricted maximum likelihood (REML).[18] Significance of the treatment effect was determined using Satterweite adjustment and confidence intervals were obtained by parametric bootstrapping[19]. Because there are some differences in age representation between the reported data cohort and the overall population (see Figure 6), weighted mixed modelling was also done using post-stratification weights at the age level.[20] Parameter estimates were very similar to the unweighted results, which is relatively unsurprising given most weights are fairly close to 1. The modelling of net effects herein uses unweighted data.

Models were also tried for a hierarchical structure of nesting class within school, however the variance accounted for by addition the school level was very low. This was also seen in plots of the variability of data across schools and classes, which indicated no obvious correlation at the school level beyond the intra-class variation. Additionally, the marginal model, where there are no random effects, was run for comparison.

Secondary data

Ministers will want to know what was achieved in terms of attendance for learners through the government initiative, and also whether this achievement would have

[16] Education counts website, Ministry of Education (https://www.educationcounts.govt.nz/directories/list-of-nz-schools).
[17] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
[18] *lme4* package available in R (authors: Douglas Bates, Martin Maechler, Ben Bolker, and Steve Walker).
[19] 500 simulations using percentile method.
[20] *WeMix* package in R (authors: Paul Bailey, Claire Kelley, Trang Nguyen, Huade Huo and Christian Kjeldsen).

occurred without HPR support. The difference the activities made would represent the 'value-add' of Ka Ora, Ka Ako. We specifically would like to know if Ka Ora, Ka Ako school recipients improved attendance rates as a result of the pilot compared with similar schools elsewhere in New Zealand that did not participate. Our approach uses the Ministry of Education's secondary datasets to match Ka Ora, Ka Ako schools with schools not receiving the support, and then track attendance statistics of the Ka Ora, Ka Ako schools and matched comparison schools over time. The difference in attendance statistics between recipient and matched schools following the start of Ka Ora, Ka Ako provides an estimate of the impact of the pilot.

A first crucial step in this interim evaluation was to determine whether it is possible to identify a group of matched schools that are similar to the school recipients in all ways except that they did not participate in the pilot. The matching is based on school-level data measured prior to the start of Ka Ora, Ka Ako (to ensure that matching variables are not affected by participation).

An important consideration for the interim evaluation is the high degree of variability in weekly attendance rates across terms and between years. This variability needs to be considered when comparing school attendance rates before and after the start of the initiative. Analysis presented on the Ministry of Education's Attendance website[21] shows that regular attendance is typically high at the beginning of a term and drops off in the last weeks of a term. This pattern is consistent across terms and years. Superimposed on this week-to-week variability is a longer-term declining trend in attendance rates. For example, regular attendance in Term 2 has dropped from 70% in 2015 to 58% in 2019. Regular attendance rates have increased post lockdown (64% with regular attendance in the last 7 weeks of Term 2, 2020). This variability will complicate the interpretation of our interim evaluation results since it will be hard to distinguish the impact of the initiative from underlying trends and post-lockdown changes.

The impact analysis was undertaken separately for Tranche 1 and Tranche 3 schools as schools started delivering the pilot in very different circumstances. Tranche 1 schools started early in the school year, prior to lockdown, whereas Tranche 3 schools started later in the school year (most started in Term 4, 2020 or later). For Tranche 1 schools, attendance performance can be tracked for several months after they started delivering the pilot, providing an opportunity to assess the longer-term impact of the pilot. By contrast, we are only able to track attendance outcomes of Tranche 3 schools for a relatively short period (about 4 to 8 weeks depending on when the schools started delivering the pilot).

Data

We used attendance data and other learner-related data sourced from the Ministry's internal databases (primarily *Learner BDS*) to undertake the matching and examine the attendance trends over time for Ka Ora, Ka Ako and non-Ka Ora, Ka Ako schools. The Ministry collects attendance data from all schools through its Term 2 Attendance Survey

[21] (https://www.educationcounts.govt.nz/data-services/national/attendanceshow)

and participation is voluntary. In addition, schools may choose to provide the Ministry with attendance information in other terms of the school year. Therefore, the attendance dataset is not complete in terms of all schools participating in the pilot nor all schools in New Zealand.

For the matching, we used weekly Term 2 attendance data prior to the start of the pilot: from 2015 to 2020 (Tranche 3 schools) or from 2015 to 2019 (Tranche 1 schools). These datasets are relatively complete (93% of all Tranche 1 and 3 schools had some Term 2 attendance records in these years). However, records for other Terms were less complete. For example, 33 of the 42 eligible Tranche 1 schools (79%) and 92 out of the 123 eligible Tranche 3 schools (75%) had Term 4, 2020 attendance data (although attendance records may not have been complete across the term). For Tranche 3 schools, we also also matched on attendance measures in the 3 months prior to the start date of the pilot and this meant that we lost some schools from the sample because they did not have Term3 or Term 4 attendance records. This analysis excluded schools with no enrolled learners in Years 1 to 8.

We also obtained school characteristics information, such as the ethnicity composition, school decile and rural/urban nature of a school from the School Directory on the Ministry of Education website (as at April 2021).

Propensity score matching and impact analysis

We used propensity score matching to identify the matched comparison group of schools with a basket of school-level variables[22]. These included:

- Prior attendance rates by type (Present, Unjustified Absence, or Justified Absence[23]) based on averaged Term 2 rates between 2015 and 2019 (Tranche 1) or from 2015 to 2020 (Tranche 3).
- Prior rates of unjustified absence in the three months leading up to the start date of the pilot (Tranche 3 only).
- Prior participation of the learner in school-based interventions (including, e.g., different types of Learner support, stand-downs, suspensions, truancy related) based on averaged rates between 2015 and 2019 (Tranche 1) or from 2015 to 2020 (Tranche 3).
- School Decile

[22] The pilot was delivered at a school-level and we did not know which students participated in the initiative. This was the main reason for selecting the school as the unit of analysis. An alternative approach might be to attempt a student-level analysis and estimate the impact on student-level attendance of being enrolled in a school that participated in the pilot. This would require matching individual students to similar students in comparable schools. Student records could only be retained for schools with nearly 100% matching of students, since we would not know whether there was bias in the unmatched student group. While the numbers in the sample would be significantly higher for student-level estimates, as opposed to school-level estimates, so too would the variability and there is the additional concern that the within-school variability might not be well modelled by the matching process.

[23] The categories Present, Justified Absence and Unjustified Absence are standard categories used by the Ministry of Education in their analysis. They are derived from more detailed attendance codes. See https://www.educationcounts.govt.nz/data-services/national/attendanceshow).

- Proportion of Māori students[24]
- Total Roll
- Whether a school was located in a major urban area
- Whether a school was classified as full primary
- Whether a school was co-educational

Other variables were considered but were dropped from the matching algorithm if they did not significantly affect the propensity to be treated (i.e. to be a recipient of Ka Ora, Ka Ako) or they were too highly correlated with other included variables. We also imposed some exact matching conditions so that schools were only matched to each other if they were in the same broad school decile group (1-3; 4-6; 7-8) and if they were classified in the same quartile of average proportion of unjustified absences in Term 2 between 2015 and 2020. The later restriction was imposed for Tranche 3 schools in order to look at the impact for particularly vulnerable schools, such as those that started with higher-than-average absence rates prior to the initiative. We were unable to impose the restriction for Tranche 1 schools because the number of participating schools was too small.

For the impact analysis, we required attendance information in the first month after they started delivering the pilot in order to measure the impact on attendance outcomes. As discussed above, this excludes schools that started too late in the year or that have no attendance data in Term 4, 2020. In addition, for Tranche 3 schools we also restricted schools if they did not have attendance records in the 3 months prior to the start of the initiative because we wanted to ensure that the average attendance performance for treated and matched comparison schools was similar before starting the pilot. Tranche 1 schools started early in the 2020 school year so they did not have prior attendance records in the same year and no restriction was imposed. However, we did check that attendance records in 2019 were similar for matched comparison and treated schools.

The start dates for each Tranche ranged across a number of weeks. For Tranche 3 schools most dates fell between weeks 41 to 45; for Tranche 1 most start dates fell between weeks 4 to 8. Comparison schools did not have a start date because they did not participate in the pilot. As part of the analysis, each potential comparison school was assigned to all the possible start weeks associated with pilot schools and a dataset was constructed of all treated schools and all possible comparison school-start week combinations. Pilot schools were matched to potential comparison schools with the same start week.

[24] Proportions of Pacific learners within schools were not included in the basket of matching variables because these proportions were small (3-4%) and roughly balanced between the pilot schools and comparison schools (c.f Table 4 in the Appendix), both before and after matching. By contrast, the proportion of Māori students before matching was high with large differences between the pilot schools and comparison schools (71% to 40%), and therefore required inclusion in the matching. It should be noted that if the pilot is implemented in different regions, it would likely be necessary to include Pacific prop in the match.

A few other restrictions were placed on the pool of comparison schools. High decile schools (9 and 10) were excluded as potential matches because they were not suitable matches for the pilot schools. Private schools were dropped for the same reason. Schools with missing information about ethnic composition or other school characteristics used in the matching process were also excluded.

Our preferred method was nearest neighbour matching with replacement. We used 10 nearest-neighbours and a caliper of 0.2 of the standard deviation in propensity scores. The choice of caliper involved a trade-off between statistical precision and a high-quality match. If the caliper was too large, then we would have included schools that were not well matched to the Ka Ora, Ka Ako schools. On the other hand, if the caliper was too small, we would have well-matched recipient schools, but only a small number. In that case, estimates will have low levels of statistical precision and a high proportion of the pilot group would not have been included in the analysis.

The last step in the matching process was to test the quality of the match. The key assumption was that after matching, there were no significant differences between the pilot schools and their matched comparison schools in terms of their likelihood of being treated. We tested this assumption by comparing the mean observed characteristics and the standardised mean differences for the two groups before and after matching. We regarded the two groups as balanced if there were no significant differences in the covariate means between the treated and comparison groups ($p<0.05$). We also visually inspected the distribution of the covariates by propensity score to ensure that the groups were balanced.

The final step was to use the weights obtained in the matching process to estimate the impact of the pilot on attendance outcomes. Outcomes were calculated by (weighted) averaging attendance measures across all Year 1 to 8 learners in the treated schools and all Year 1 to 8 learners in the matched comparison schools across 4-week periods following the start of the initiative, e.g. starting with the week associated with the pilot start date (week 0) up to week 3, followed by week 4 to week 7, etc. Records from weeks associated with low numbers of school returns (e.g. week 4, 50 and 51 and school holiday weeks) were excluded when calculating outcomes. Similarly, attendance records associated with the national lockdown weeks were also excluded from the impact analysis. For the Tranche 3 analysis we also excluded Auckland schools to avoid complications associated with the Auckland lockdowns in Term 3[25]. Therefore, there are some 4-week periods following pilot start weeks that were not included in the impact analysis results. Standard errors and confidence intervals were estimated by bootstrapping (N= 500) in order to account for the uncertainty associated with the first

[25]We retained Auckland region schools in the comparison group for Tranche 1 schools as the Auckland lockdown occurred in the later part of the follow-up period and was therefore not as big a concern as for Tranche 3 schools that started in Term 4 closer to the lockdown period.

stage of the matching process (estimating the likelihood of being treated)[26]. We used the R package *Matchit*[27] for this analysis.

Matching Results

The following tables present mean proportions or means averaged across all students aged 5 to 13 and all schools (by treatment status, i.e. participating pilot schools or matched comparison schools). The first table (Table 4) refers to Tranche 3 schools and comparison schools and the second table refers to Tranche 1 schools and comparison schools (Table 5). The tables demonstrate that participating schools differed from non-participating schools and how matching brought the groups closer together, at least for observed variables shown in the tables. The results before matching are shown in the right-hand panel and the results after matching are shown in left hand panel. For example, before matching the average proportion of Māori students in the potential group of matches for Tranche 3 schools (the non-participating schools) was 40% whereas after matching the proportion was the same as the treated schools at 68%. In another example, the mean proportion of time classified as present in Term 2 across all schools between 2015 and 2020 was 88% for Tranche 3 Ka Ora, Ka Ako schools and 90% for potential comparison schools. After matching both groups had the same rates at 88%.

Table 4: Characteristics relevant to matched and unmatched schools- Tranche 3 and Expansion schools

	After matching		Before matching	
School-level Variables	Matched comparison group	Tranche 3 schools	Non-participating schools	Tranche 3 schools
Demographic				
Proportion of Māori students	68%	68%	40%	71%
Proportion of European students	21%	22%	44%	20%
Proportion of Asian students	4%	4%	6%	3%
Proportion of MELAA students	1%	1%	2%	1%
Proportion of Pasifika student	6%	5%	4%	3%
Proportion of schools located in main urban area	53%	46%	40%	48%
Proportion of schools that are coed	100%	100%	100%	100%
Total_roll	192	204	182	189
Proportion of schools classified as full primary school	44%	40%	52%	45%
School Decile				

[26] We also compared the bootstrapped errors to robust standard errors that account for the fact that comparison schools might be used more than once in the matched dataset. The bootstrapped errors were similar or slightly larger than the robust standard errors.

[27] Ho DE, Imai K, King G, Stuart EA (2011). "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, **42**(8), 1–28. https://www.jstatsoft.org/v42/i08/.

School-level Variables	After matching		Before matching	
	Matched comparison group	Tranche 3 schools	Non-participating schools	Tranche 3 schools
1 to 3	77%	77%	47%	81%
4 to 6	23%	23%	53%	19%
7 to 10	0%	0%	0%	0%
Regional Council				
Bay of Plenty Region	2%	42%	5%	43%
Canterbury Region	4%	0%	11%	0%
Gisborne Region	1%	12%	2%	12%
Hawke's Bay Region	4%	32%	4%	33%
Manawatu-Wanganui Region	15%	0%	14%	0%
Northland Region	27%	0%	13%	0%
Taranaki Region	5%	0%	6%	0%
Waikato Region	24%	14%	19%	11%
Wellington Region	14%	0%	11%	0%
Other	4%	0%	15%	0%
Attendance indicators (2015-2020 Term 2)				
Proportion classified as Present	88%	88%	90%	88%
Proportion classified as Justified Absence	6%	6%	6%	6%
Proportion classified as Unjustified Absence	6%	5%	4%	5%
Proportion classified as Unjustified Absence (by Quartile)				
1 (lowest % unjustified absence)	11%	11%	29%	18%
2	18%	18%	28%	16%
3	37%	37%	24%	28%
4 (highest % unjustified absence)	35%	35%	20%	39%
Behavioural indicators				
Proportion of students with prior suspensions	0.7%	0.7%	0.8%	0.6%
Proportion of students with prior standdowns	5%	4%	6%	4%
Proportion of students with prior NETS	6%	5%	4%	6%
Proportion of students with prior truancy issues	4%	4%	2%	4%
Proportion of students with a prior enrolment in Learning Support (prev. Special Education)	12%	10%	11%	10%
Proportion of students enrolled in prior intervention of any type	44%	41%	41%	40%
Number of schools in group	**202**	**57**	**699**	**69**
% Matched		**83%**		

Table 5: Characteristics relevant to matched and unmatched schools- Tranche 1 schools

School-level Variables	After matching		Before matching	
	Matched comparison group	Tranche 1 schools	Non-participating schools	Tranche 1 schools
Demographic				
Proportion of Māori students	78%	80%	39%	83%
Proportion of European students	11%	11%	37%	9%
Proportion of Asian students	3%	2%	8%	1%
Proportion of MELAA students	1%	0%	2%	0%
Proportion of Pasifika student	7%	6%	12%	6%
Proportion of schools located in main urban area	60%	65%	50%	73%
Proportion of schools that are coed	100%	100%	100%	100%
Total_roll	149	259	214	236
Proportion of schools classified as full primary school	46%	31%	48%	33%
School Decile				
1 to 3	96%	96%	47%	97%
4 to 6	4%	4%	53%	3%
7 to 10	0%	0%	0%	0%
Regional Council				
Auckland Region	9%		22%	
Bay of Plenty Region	2%	65%	4%	54%
Canterbury Region	5%		8%	
Gisborne Region	1%	8%	1%	19%
Hawke's Bay Region		27%	3%	27%
Manawatu-Wanganui Region	15%		11%	
Northland Region	28%		10%	
Taranaki Region	6%		5%	
Waikato Region	19%		15%	
Wellington Region	13%		8%	
Other	3%		13%	
Attendance indicators (2015-2020 Term 2)				
Proportion classified as Present	88%	87%	90%	87%
Proportion classified as Justified Absence	6%	7%	6%	6%
Proportion classified as Unjustified Absence	6%	6%	4%	6%
Proportion classified as Unjustified Absence (by Quartile)				
1 (lowest % unjustified absence)	8%	12%	26%	11%
2	18%	8%	26%	8%
3	28%	35%	25%	27%
4 (highest % unjustified absence)	46%	46%	23%	54%
Behavioural indicators				
Proportion of students with prior suspensions	1%	1%	1%	1%
Proportion of students with prior standdowns	6%	5%	5%	6%

School-level Variables	After matching		Before matching	
	Matched comparison group	Tranche 1 schools	Non-participating schools	Tranche 1 schools
Proportion of students with prior NETS	6%	6%	4%	6%
Proportion of students with prior truancy issues	5%	4%	3%	5%
Proportion of students with a prior enrolment in Learning Support (prev. Special Education)	11%	11%	10%	12%
Proportion of students enrolled in prior intervention of any type	43%	40%	44%	42%
Number of schools in group	**134**	**26**	**1046**	**33**
% Matched		**79%**		

All the key variables for the impact assessment (particularly the prior attendance and behavioural indicators and socio-economic indicators such as Decile) were statistically equivalent for the treated schools and the matched comparison schools.

The table also shows the Regional Council distribution of Ka Ora, Ka Ako schools and comparison schools. We did not expect those to be the same after matching because we needed to draw schools from other regions to act as comparators. Over half the matched comparison schools were from the Waikato, Northland and Manawatu-Wanganui Regional Councils.

A majority (83%) of the 69 Tranche 3 participating schools were matched. This means that any impact estimate can be considered representative of the treated group. A low proportion of potential comparison schools were matched (23%) but that is expected as most of the schools in the potential pool had quite different characteristics to the participating schools.

Unfortunately, there were only a small number of Tranche 1 schools that could be used in the analysis (Table 5). For Tranche 1, 33 schools had the required attendance data and of those only 26 schools were matched (79%). The effect size would need to be large in order to discern a statistically significant effect with numbers as low as this. The table shows that differences remained after matching however most of the differences were not statistically significant, including all the variables relating to attendance and behavioural indicators and socio-economic indicators.

The adequacy of the match is presented visually in the following plots. The first plot (Figure 7) shows the average proportion of time classified as unjustified absence in 2020 for Tranche 3 schools and for potential comparison schools before matching. The second plot (Figure 8) shows the same measures after matching. The period displayed is restricted to 2020 and starts at the start of lockdown (week 15) to show the variabililty in attendance in the months leading up to Tranche 3 start weeks (weeks 41 to 45).

Figure 7: Average proportion of time classified as unjustified absence in 2020 (Tranche 3 schools and comparators), BEFORE matching

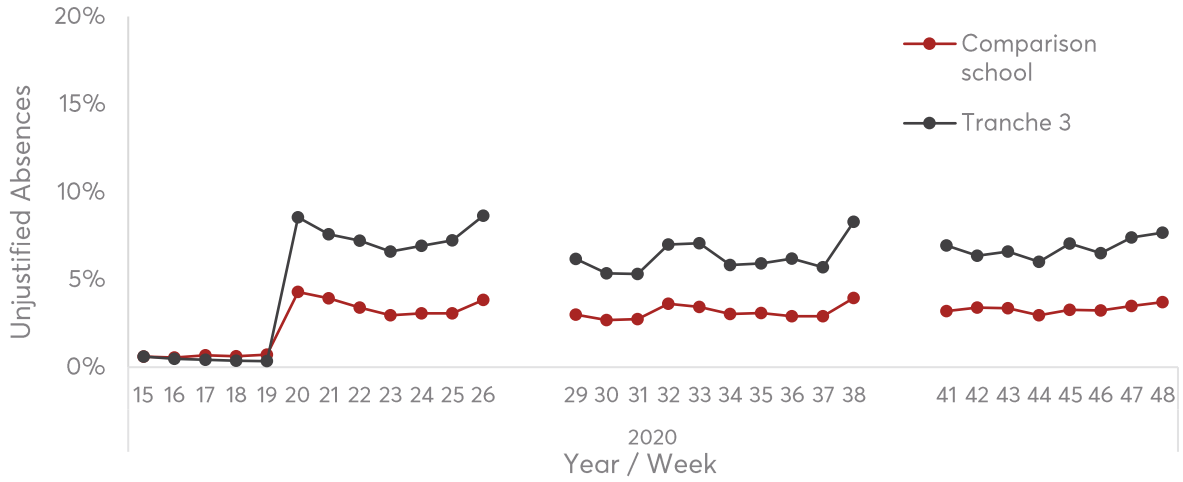


Figure 8: Average proportion of time classified as unjustified absence in 2020 (Tranche 3 and Expansion schools and comparators), AFTER matching

Before matching, the unadjusted differences in the average proportion of time classified as unjustified absence were typically between +2 to +5 percentage points higher for Tranche 3 schools compared with potential comparison schools not delivering the pilot. There were some exceptions when attendance measures were similar for both groups, particularly around lockdown (weeks 15 -19) during which unjustified absences were low for all schools. The second figure shows these differences were reduced following the matching. There was still considerable variability in week to week attendance during the year and differences between the groups particularly in the weeks following lockdown. However, the constant offset that was evident between the participating schools and comparison groups in Figure 7 was no longer present after matching.

The variable start dates between weeks 41 and 45 complicate the interpretation of these plots. At the start of Term 4, some schools had already started receiving the pilot but other schools wouldn't start for up to 4 more weeks. Therefore, if there was an immediate impact this could be obscured by averaging across calendar weeks. For this reason, all measures used in the matching and impact estimates were based on the week/month relative to the start date. The next table (Table 6) is presented to emphasise that there were no statistically significant differences in the attendance indicators used in the impact estimates between the treated and matched comparison schools in the months leading up to the start of the pilot. This is important to justify the difference-in-difference estimates that are presented later. The table shows monthly averages in 4-week periods up from 24 weeks prior to the start week (week -24) up to the week immediately prior to the start week (week -1). The significance level of a t-test in the difference of the means is shown. Before matching, most of the difference in means were statistically significantly at the 5% level, after matching the differences were not significant.

Table 6: Prior monthly attendance measures for Tranche 3 schools measured from 24 weeks before the start date, up to the week immediately prior

	AFTER MATCHING			BEFORE MATCHING		
Attendance indicators in monthlys PRIOR to start	Matched comparison group	Tranche 3 schools	Significance level	Non-participating schools	Tranche 3 schools	Significance level
Weeks -4 to -1						
Proportion classified as Present	86%	86%	0.96	90%	86%	0.00
Proportion classified as Justified Absence	6%	7%	0.65	5%	7%	0.00
Proportion classified as Unjustified Absence	8%	8%	0.59	5%	7%	0.00
Weeks -8 to -5						
Proportion classified as Present	85%	86%	0.87	88%	86%	0.01
Proportion classified as Justified Absence	7%	7%	0.57	6%	7%	0.01
Proportion classified as Unjustified Absence	7%	7%	0.68	6%	7%	0.16
Weeks -12 to -9						
Proportion classified as Present	86%	87%	0.86	89%	87%	0.00
Proportion classified as Justified Absence	7%	7%	0.66	7%	7%	0.06
Proportion classified as Unjustified Absence	7%	6%	0.44	5%	6%	0.01
Weeks -16 to -13						

	AFTER MATCHING			BEFORE MATCHING		
Attendance indicators in monthlys PRIOR to start	Matched comparison group	Tranche 3 schools	Significance level	Non-participating schools	Tranche 3 schools	Significance level
Proportion classified as Present	81%	83%	0.63	89%	84%	0.00
Proportion classified as Justified Absence	9%	8%	0.69	6%	8%	0.00
Proportion classified as Unjustified Absence	8%	8%	0.77	5%	8%	0.00
Weeks -20 to -17						
Proportion classified as Present	84%	84%	0.89	87%	85%	0.00
Proportion classified as Justified Absence	7%	8%	0.38	8%	8%	0.96
Proportion classified as Unjustified Absence	8%	8%	0.78	5%	8%	0.00
Weeks -24 to -21						
Proportion classified as Present	91%	94%	0.70	91%	93%	0.00
Proportion classified as Justified Absence	3%	2%	0.24	4%	3%	0.00
Proportion classified as Unjustified Absence	5%	4%	0.51	5%	4%	0.08

For example, note the close match in the proportion of time classified as unjustified absence for weeks immediately prior to most start dates (week -4 to week -1). The average attendance measures averaged across the 4 weeks immediately prior to the start date are also shown in Table 6. Before matching, the average proportion of time classified as unjustified absence in the 4 weeks prior to starting the initiative was 7% for treated, compared to 5% for potential comparison schools. After matching there were no differences in this measure in the 4 weeks immediately preceding the start of participation (both 8%). Therefore, if the average attendance measures differ for the treated and matched comparison schools after the start date, we can be more confident that this was the impact due to the initiative rather than any pre-existing differences in attendance.

The next two figures show that quality of the match for Tranche 1 schools which mainly started early in the 2020 school year. The plots demonstrate that the average proportion of time classified as unjustified absence was consistently higher by about 3 to 6 percentage points for Tranche 1 schools compared to potential comparison schools before matching. After matching, the differences were reduced to close to zero prior to the start date (i.e. in 2019). The plots show that the proportion of time spent classified as unjustified absence rates was lower for Tranche 1 schools following the start of the initiative (early 2020) and this difference persisted for several weeks and was apparent post-lockdown. We will assess whether this difference was statistically significant in the next section where impact estimates are presented.

Figure 9: Average proportion of time classified as unjustified absence in 2020 (Tranche 1 schools and comparators), BEFORE matching

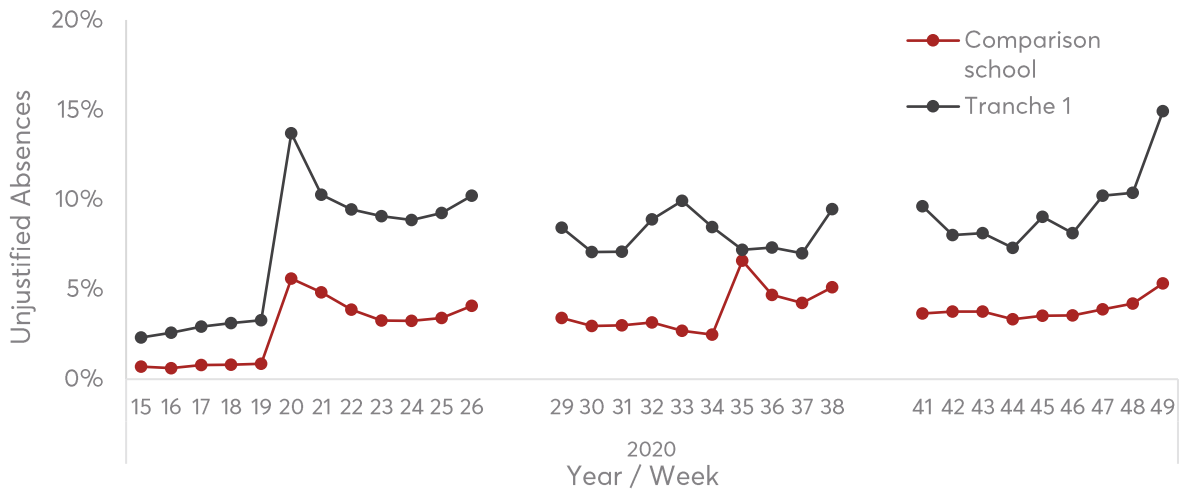


Figure 10: Average proportion of time classified as unjustified absence in 2020 (Tranche 1 schools and comparators), AFTER matching

Impact estimates

Table 7 summarises the impact of the pilot on attendance measures for Tranche 3 schools. Three different types of attendance outcomes are presented:

1) The proportion of time spent classified as Present
2) The proportion of time spent classified as Justified Absence
3) The proportion of time spent classified as Unjustified Absence

The outcomes are measured in 4-weekly periods following the start of delivery of the pilot to assess the time-dependent nature of the impact. For each outcome type and outcome period, the table shows the average value of the outcome for the matched

comparison group and pilot schools. The third column presents the impact estimates (the average treatment effect on the treated). The standard errors and 95 % confidence intervals are also shown.

Table 7: Impact estimates on attendance outcomes for Tranche 3 and expansion schools

Averaged 4-week attendance outcomes following the start of pilot delivery	Mean Comparison	Mean Tranche 3	Impact estimate	SE	95% CIs	
Proportion of time as Present						
Week 0 to Week 3 after start	86.9%	87.1%	0.2%	0.8%	-1.4%	1.8%
Week 4 to Week 7 after start	84.6%	83.6%	-1.0%	1.1%	-3.2%	1.2%
Proportion of time as Justified Absence						
Week 0 to Week 3 after start	5.8%	5.9%	0.1%	0.5%	-0.9%	1.2%
Week 4 to Week 7 after start	6.9%	7.1%	0.2%	0.8%	-1.3%	1.8%
Proportion of time as Unjustified Absence						
Week 0 to Week 3 after start	7.3%	7.0%	-0.3%	0.7%	-1.6%	1.0%
Week 4 to Week 7 after start	8.5%	9.3%	0.8%	0.9%	-1.0%	2.6%

The table shows there was no significant impact on attendance outcomes for Tranche 3 for up to 8 weeks following the start of the initiative. All impact estimates were close to 0 and all confidence intervals show that the estimates were not significantly different than 0. For example, the proportion of time spent classified as unjustified absence for matched comparison schools in the 4 weeks following the start of the programme was 7.3% compared to 7.0% for Tranche 3 schools, which appears to be a slightly positive impact due to the pilot of – 0.3 percentage points. However, the error is 0.7 percentage points, so the 95% confidence intervals fall between -1.6 percentage points and 1.0 percentage points, i.e. there was no significant impact. The results show there was no immediate effect on attendance outcomes, at least not when averaged across all the Year 1 to 8 learners. However, this does not mean there was no impact on the attendance patterns of individual students in participating pilot schools or even on attendance patterns for specific classes or Year levels. If we had more nuanced participation information, e.g. we knew which students or which classes participated in the pilot, then we could have undertaken a more focused impact estimate rather than averaging across all learners at a school.

We also calculated the impact for schools with higher than average absence rates prior to the pilot. To do this, we calculated the impact for schools that were in the top 2 quartiles of unjustified absence rates. The impact estimates for this sub-group of schools were also statistically insignificant (results not shown).

It is possible that attendance patterns may take some time to change following participation in the initiative. We were only able to track attendance in Tranche 3 schools for a few weeks, however it would be worth undertaking a similar analysis for these schools again once more time has elapsed to determine whether there is a lagged effect on attendance patterns.

We can examine the longer-term impact on attendance by looking at the impact estimates for Tranche 1 schools. These are shown in the following table for the same types of attendance outcomes as shown for Tranche 3 schools, but with longer follow-up periods.

Table 8: Impact estimates on attendance outcomes for Tranche 1 schools

Averaged 4-week attendance outcomes following the start of pilot delivery	Mean Comparison	Mean Tranche 1	Impact estimate	SE	95% CIs	
Proportion of time as Present						
Week 0 to Week 3 after start	87.1%	86.8%	**-0.4%**	2.4%	-5.1%	4.3%
Week 4 to Week 7 after start	81.1%	77.6%	**-3.6%**	4.2%	-11.8%	4.6%
Week 16 to Week 19 after start	82.3%	82.4%	**0.1%**	2.5%	-4.8%	5.0%
Week 20 to Week 23 after start	81.2%	82.7%	**1.5%**	1.7%	-1.8%	4.8%
Week 24 to Week 27 after start	84.6%	84.9%	**0.3%**	1.7%	-3.0%	3.6%
Week 28 to Week 31 after start	84.2%	83.8%	**-0.4%**	1.7%	-3.7%	2.9%
Week 36 to Week 39 after start	85.4%	85.5%	**0.1%**	1.3%	-2.4%	2.6%
Week 40 to Week 43 after start	82.8%	82.6%	**-0.2%**	1.6%	-3.3%	2.9%
Proportion of time as Justified Absence						
Week 0 to Week 3 after start	5.7%	6.9%	**1.2%**	1.9%	-2.5%	4.9%
Week 4 to Week 7 after start	15.2%	18.4%	**3.2%**	4.2%	-5.0%	11.4%
Week 16 to Week 19 after start	9.9%	9.3%	**-0.5%**	1.8%	-4.0%	3.0%
Week 20 to Week 23 after start	11.4%	9.8%	**-1.6%**	2.4%	-6.3%	3.1%
Week 24 to Week 27 after start	9.5%	8.3%	**-1.2%**	1.6%	-4.3%	1.9%
Week 28 to Week 31 after start	9.5%	9.1%	**-0.4%**	1.6%	-3.5%	2.7%
Week 36 to Week 39 after start	8.1%	7.0%	**-1.0%**	1.4%	-3.7%	1.7%
Week 40 to Week 43 after start	8.7%	8.1%	**-0.6%**	1.8%	-4.1%	2.9%
Proportion of time as Unjustified Absence						
Week 0 to Week 3 after start	7.9%	6.4%	**-1.5%**	1.4%	-4.2%	1.2%
Week 4 to Week 7 after start	8.1%	8.0%	**-0.1%**	2.5%	-5.0%	4.8%
Week 16 to Week 19 after start	13.4%	10.5%	**-2.9%**	2.7%	-8.2%	2.4%
Week 20 to Week 23 after start	12.2%	10.1%	**-2.1%**	2.5%	-7.0%	2.8%
Week 24 to Week 27 after start	10.3%	8.7%	**-1.6%**	2.0%	-5.5%	2.3%
Week 28 to Week 31 after start	11.4%	9.5%	**-1.9%**	2.0%	-5.8%	2.0%
Week 36 to Week 39 after start	11.2%	8.9%	**-2.4%**	1.7%	-5.7%	0.9%

The results show that there were no statistically significant impacts on attendance patterns for Tranche 1 schools, even up to 43 weeks following the start of delivery of the pilot. The average proportion of time classified as Unjustified absence for Tranche 1 schools was around 2 percentage points lower than for matched comparison schools for most follow-up periods. However, the associated standard errors were high, presumably associated with the small number of schools, and hence the impact estimates were not statistically significant. A larger group of schools is needed to assess whether there has been any improvement in attendance patterns over the longer term

School staff focus-group interviews

All school leaders involved in the evaluation were asked to take part in the focus group interviews. These school leaders were asked to invite school staff to take part in a zoom session, facilitated by two evaluators over a 30-minute period. The brief focus group discussion allowed for both participant-generated themes, as well as allowed for exploring the themes relevant to the programme theory of change and key evaluation questions (c.f. Appendix E).

The invited schools included both treatment and control group schools. The inclusion of all schools allowed the evaluation team to explore potential confounding variables associated with any detected changes. The semi-structured interview approach explored any detected change with the group, examining both the order of events ("Modus Operandi Method") and alternative causal mechansims ("General Elimination Method"). Such techniques are useful within impact-focussed evaluations.

Focus group interviews were conducted with nine different schools, including six schools that had been delivering the Ka Ora, Ka Ako programme and three that had not yet been delivering the programme.

Provider self-assessments

Provider's self-assessments were used alongside the school lunch photos (see below) to estimate the percentage of lunch options meeting Ministry of Health nutrition standards, based on 75% of the lunch made up of "green" elements. These were conducted in November only because no providers were delivering lunches in our schools during the first (baseline) data collection point.

The lunch providers to the 26 treatment schools were asked to complete a school lunch self-assessment covering lunches across a five-day period. The lunch assessment will assess the component parts of the lunch (i.e. ingredients) against the Ministry of Health guidance document, determining if 75% of the lunch is made up of healthy "green" ingredients. These weekly menus were requested from providers for the week of 16 November.

All providers were invited to take part given the small numbers that were involved[28]. Seven providers submitted the forms, representing 12 schools. Given the low response and likely bias in the responses, these data are not presented in the interim evaluation report. Nonetheless, the results align with the findings presented. Notably, when comparing whether or not the learners had any (of the 15 assessed) vegetables in their lunch or not (learner food diaries vs. lunch self-assessments), the results again showed the Ka Ora Ka Ako lunches were significantly more likely to include vegetables; notably the lunches provided by providers were 68 times more likely to include vegetables than learner lunches before the programme.

[28] Sampling is less efficient when used on small populations whereby achieving the same margin of error requires significantly higher proportional makeup.

Lunch photos

In addition to taking part in the learner food diares, the schools were asked to take pictures of a sample of school lunches, once in September (prior to Ka Ora, Ka Ako being provided) and again in November. The photos were used to get an indication of the percentage of learners consuming healthy lunch options, based on consumption of vegetables vs. packaged snacks and sweet items. The lunch photos also provided an opportunity to validate the availability of food options according to that reported in the learner food diaries.

All schools invited to take part in the evaluation (n=45) were asked to provide photos of school lunches in one class and on one day in September at the same time the food diaries were being administered. Schools submitted these either by printing and posting them, or by emailing the photos to the evaluation team.

Of the schools that were asked to participate, 18 submitted 168 separate school lunch photos at baseline, a 40% response rate. For the follow up 17 schools submitted 121 separate lunch photos (37% response rate), however after removing duplicate photos there were a total of 76. Of these 76, 27 were lunches provided by the schools through the programme and 45 were brought from home., The results were not used generalise the quality of the school lunches, but rather used alongside the food diaries to contextualise the results.

Lunch photos were assessed for availability of different food items, like that of the learner food diarys, however this was not a reliable measure of assessment as it was often not explicitly clear what items were in the lunches. A simpler method of estimating the quantities regarding the fullness of a 'lunchbox' was adopted (0=none, 1=a little, 2=less than half, 3=more than half, 4=all) for vegetables, snacks, fresh fruit.

Using the method described above two independent judges agreed on 82% of the (n=134) lunches at baseline, and 79% at follow-up, demonstrating a reasonable reliability rate.

Focus groups

School staff focus groups were used to complement the data above, further testing the evidence and alternative factors contributing to any realised change and used to answer the second evaluation question, which focused on the broader benefits for schools, their learners and whānau.

We invited staff from schools involved in the learner food diaries to take part in a focus group discussions. Providers were also asked to take part, but only three responded to the request. In total, 9 schools took part in the discussions, where we explored the changes experienced over the preceding three months in relation to food security, wellbeing and attendance, and any associated contributions to these changes. The questions were broad to allow participants the opportunity to consider any change and in light of the school context and the range of initiatives underway there (not simply the School Lunch pilot). We also used the opportunity to examine the order of events in

relation to reported change, and to test other factors that may have contributed to any change.

The 9 schools were made up of 6 schools that were involved in the pilot and 3 schools that were not yet involved by the time of the focus group. The approach, comparing any recognised changes between these two groups, allowed us to further test any reported relationships between changes and the pilot.

It was clear that there were insufficient numbers of respondents to achieve saturation, as largely new information was introduced in each of these focus groups. What was interesting is the range of factors relevant to the pilot, such as lunch provider and learner characteristics (e.g. size, energy levels), as well as other initiatives at the school appeared to influence the responses. Although we were not able to collect enough data to estimate how such factors influenced the results, these would be worthy for further exploration in any subsequent analysis.

All schools were provided with an information sheet and consent form prior to the engagement. During the engagement, all individuals were invited to ask questions about the evaluation and were asked to confirm their consent to take part. In response to the COVID pandemic, all interviews were conducted remotely through video communication platforms (e.g. Zoom).

Appendix B: Learner Food Diary results (Cross-sectional analysis)

The following tables present the results from the full populations of respondents (those in the pilot and those not in the pilot) in the September surveys and the November surveys. The results are presented both as unweighted and weighted data, with the weighted data weighted at the level of respondent age to reflect the overall population of learners. Confidence intervals (95%) are obtained by robust estimates to account for the classroom-level clustering that was part of the sampling design[29]. The results provide a generalisable statistics on the overall prevalence of food availability, consumption, satiety and wellbeing across primary and intermediate learners in the Hawkes Bay and Bay of Plenty.

[29] *sandwich* package available in R (authors: Zeileis A, Köll S, Graham N).

Table 9: Numbers of learners lunches where different food items were available (unweighted) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

Available (unweighted) Food Item	Week 1 (September)														Week 2 (November)												
	Control							Treatment							Control							Treatment						
	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper
Tomato	5.6%	6.0%	6.3%	8.0%	**6.5%**	5.4%	7.6%	7.5%	7.2%	6.2%	6.0%	**6.8%**	6.2%	7.3%	4.9%	7.1%	5.6%	10.7%	**6.8%**	5.6%	8.1%	18.5%	18.9%	19.0%	27.3%	**20.8%**	19.7%	21.8%
Pumpkin	0.8%	0.2%	1.0%	1.0%	**0.8%**	0.4%	1.1%	1.8%	1.2%	1.1%	0.8%	**1.2%**	1.0%	1.5%	0.7%	0.5%	0.2%	1.6%	**0.7%**	0.3%	1.1%	0.5%	0.7%	2.1%	0.8%	**1.1%**	0.8%	1.3%
Potato	3.5%	2.9%	3.4%	3.5%	**3.3%**	2.5%	4.1%	5.5%	4.8%	3.1%	2.8%	**4.1%**	3.7%	4.6%	3.4%	3.0%	2.8%	6.0%	**3.6%**	2.7%	4.5%	19.3%	5.0%	12.0%	8.4%	**11.3%**	10.5%	12.1%
Peas	1.4%	2.3%	1.4%	2.3%	**1.9%**	1.3%	2.5%	2.4%	2.2%	1.5%	1.1%	**1.8%**	1.5%	2.1%	1.6%	1.6%	1.6%	2.5%	**1.8%**	1.1%	2.4%	12.0%	7.3%	12.6%	2.7%	**8.7%**	8.0%	9.5%
Lettuce	9.9%	10.5%	11.5%	11.3%	**10.8%**	9.4%	12.2%	8.0%	6.5%	7.4%	7.1%	**7.2%**	6.7%	7.8%	9.2%	9.2%	10.6%	10.7%	**9.8%**	8.4%	11.3%	29.4%	26.6%	26.4%	34.6%	**29.2%**	28.0%	30.3%
Kumara	1.2%	0.8%	1.0%	1.6%	**1.2%**	0.7%	1.6%	1.8%	1.9%	1.1%	0.9%	**1.4%**	1.2%	1.7%	1.1%	0.7%	0.7%	1.9%	**1.0%**	0.6%	1.5%	1.0%	1.5%	3.2%	0.8%	**1.6%**	1.3%	2.0%
Green leafy veg	1.4%	0.8%	1.4%	1.6%	**1.3%**	0.8%	1.8%	1.7%	1.2%	1.5%	1.1%	**1.4%**	1.1%	1.6%	2.0%	1.1%	0.9%	0.9%	**1.3%**	0.7%	1.8%	11.8%	5.7%	4.8%	8.8%	**7.8%**	7.1%	8.5%
Cucumber	4.7%	5.8%	7.3%	5.3%	**5.8%**	4.8%	6.9%	6.6%	7.0%	7.7%	7.2%	**7.1%**	6.5%	7.7%	7.4%	6.7%	8.2%	6.6%	**7.3%**	6.0%	8.5%	12.0%	19.0%	19.1%	24.8%	**18.5%**	17.5%	19.4%
Corn	2.1%	2.5%	1.6%	3.1%	**2.3%**	1.6%	3.0%	4.2%	3.7%	3.0%	3.0%	**3.5%**	3.1%	3.9%	1.3%	1.8%	2.1%	4.1%	**2.2%**	1.5%	2.9%	18.2%	11.5%	17.0%	4.7%	**13.1%**	12.2%	13.9%
Cauliflower	0.4%	0.4%	0.2%	0.0%	**0.3%**	0.0%	0.5%	1.1%	0.8%	0.8%	1.2%	**1.0%**	0.8%	1.2%	0.4%	0.7%	0.5%	0.6%	**0.6%**	0.2%	0.9%	1.6%	5.5%	0.8%	0.4%	**2.2%**	1.8%	2.5%
Carrot	16.5%	14.3%	14.1%	11.9%	**14.2%**	12.7%	15.7%	13.8%	14.0%	13.5%	12.7%	**13.5%**	12.8%	14.3%	13.4%	12.0%	12.7%	14.2%	**13.0%**	11.3%	14.6%	48.9%	37.1%	35.4%	38.9%	**40.2%**	39.0%	41.5%
Capsicum	1.4%	1.9%	2.6%	1.4%	**1.9%**	1.3%	2.5%	2.7%	2.0%	2.0%	2.0%	**2.2%**	1.9%	2.5%	1.1%	1.6%	2.1%	2.5%	**1.8%**	1.1%	2.4%	8.6%	11.3%	5.2%	9.0%	**8.6%**	7.9%	9.3%
Cabbage	3.1%	1.9%	3.0%	2.5%	**2.6%**	1.9%	3.3%	2.9%	2.6%	2.2%	1.8%	**2.4%**	2.1%	2.7%	3.6%	2.1%	1.9%	2.5%	**2.5%**	1.8%	3.3%	9.2%	13.7%	8.7%	11.7%	**10.8%**	10.0%	11.6%
Brocolli	1.2%	1.4%	1.0%	1.6%	**1.3%**	0.8%	1.8%	2.5%	1.8%	1.9%	1.6%	**2.0%**	1.7%	2.3%	1.3%	1.1%	0.2%	1.9%	**1.1%**	0.6%	1.6%	3.6%	7.5%	1.5%	1.3%	**3.6%**	3.1%	4.1%
Beans	0.8%	0.4%	0.8%	0.6%	**0.7%**	0.3%	1.0%	1.9%	1.6%	1.3%	1.1%	**1.5%**	1.2%	1.8%	0.9%	0.5%	0.7%	1.3%	**0.8%**	0.4%	1.2%	1.9%	1.1%	3.1%	2.0%	**2.0%**	1.6%	2.3%
Sugar coated snacks	4.1%	3.5%	4.8%	7.2%	**4.9%**	3.9%	5.8%	7.0%	6.3%	5.8%	6.3%	**6.4%**	5.8%	6.9%	4.7%	3.9%	6.1%	3.1%	**4.6%**	3.5%	5.6%	3.1%	2.8%	2.6%	2.2%	**2.7%**	2.3%	3.1%
Popcorn	8.9%	7.4%	9.7%	10.3%	**9.0%**	7.8%	10.3%	11.4%	13.5%	12.0%	11.6%	**12.1%**	11.4%	12.9%	10.3%	10.8%	12.9%	11.6%	**11.4%**	9.8%	12.9%	12.6%	7.1%	10.1%	10.2%	**10.0%**	9.2%	10.7%
Muesli/fruit bar	37.1%	38.6%	37.0%	38.9%	**37.9%**	35.8%	40.0%	38.1%	37.1%	36.8%	37.7%	**37.4%**	36.3%	38.5%	41.6%	40.5%	41.6%	43.4%	**41.7%**	39.3%	44.1%	18.0%	19.5%	16.4%	16.5%	**17.7%**	16.7%	18.6%
Lollies	15.5%	15.1%	20.4%	17.5%	**17.1%**	15.5%	18.8%	19.0%	19.2%	17.1%	17.4%	**18.2%**	17.4%	19.1%	16.8%	14.0%	16.5%	11.6%	**15.0%**	13.2%	16.7%	9.5%	9.4%	8.3%	7.8%	**8.8%**	8.1%	9.5%
Cracker	22.9%	23.6%	23.2%	18.3%	**22.0%**	20.2%	23.9%	24.9%	24.7%	23.5%	22.6%	**24.0%**	23.0%	24.9%	25.3%	24.1%	21.2%	22.0%	**23.3%**	21.2%	25.3%	15.6%	18.4%	12.1%	13.9%	**15.1%**	14.2%	16.0%
Cornchips	12.2%	15.3%	18.0%	18.5%	**16.0%**	14.4%	17.6%	12.9%	11.4%	11.5%	12.6%	**12.1%**	11.3%	12.8%	17.4%	14.9%	13.4%	17.0%	**15.6%**	13.9%	17.4%	13.9%	11.2%	16.1%	14.3%	**13.8%**	12.9%	14.7%
Cookies	41.6%	43.0%	45.7%	42.8%	**43.3%**	41.1%	45.5%	42.8%	42.7%	40.5%	41.6%	**42.0%**	40.8%	43.1%	41.4%	39.5%	39.8%	40.3%	**40.2%**	37.9%	42.6%	24.2%	20.3%	22.9%	18.8%	**21.6%**	20.6%	22.7%
Chocolate	10.9%	8.9%	9.9%	14.4%	**11.0%**	9.6%	12.4%	12.4%	11.8%	11.0%	10.2%	**11.4%**	10.7%	12.1%	9.6%	10.3%	10.1%	9.4%	**9.9%**	8.5%	11.4%	7.8%	7.8%	6.4%	6.0%	**7.0%**	6.4%	7.7%
Potato chips/chippies	56.7%	54.5%	60.6%	60.1%	**57.9%**	55.8%	60.1%	64.6%	64.0%	62.2%	62.5%	**63.4%**	62.3%	64.5%	57.9%	57.7%	55.8%	59.1%	**57.5%**	55.1%	59.9%	35.7%	32.4%	32.2%	32.2%	**33.2%**	32.0%	34.4%
Cake of muffin	13.6%	14.9%	15.2%	13.2%	**14.3%**	12.7%	15.8%	15.0%	13.6%	14.9%	13.7%	**14.4%**	13.6%	15.1%	13.6%	12.4%	15.1%	11.6%	**13.3%**	11.6%	14.9%	25.6%	22.6%	22.9%	28.5%	**24.8%**	23.7%	25.9%
Watermelon	3.9%	2.7%	2.2%	3.7%	**3.1%**	2.3%	3.9%	5.0%	4.3%	3.6%	3.5%	**4.1%**	3.7%	4.6%	2.7%	2.8%	3.3%	1.9%	**2.7%**	1.9%	3.5%	4.6%	1.7%	1.5%	2.8%	**2.7%**	2.3%	3.1%
Stonefruit (peach/plums)	2.5%	4.8%	2.0%	2.7%	**3.0%**	2.3%	3.8%	3.6%	3.4%	2.9%	2.8%	**3.2%**	2.8%	3.6%	4.9%	1.6%	3.3%	3.1%	**3.3%**	2.4%	4.1%	6.4%	4.7%	6.3%	11.0%	**7.0%**	6.3%	7.6%
Pineapple	2.9%	2.3%	2.4%	2.7%	**2.6%**	1.9%	3.3%	3.7%	3.9%	4.1%	3.4%	**3.8%**	3.4%	4.2%	4.3%	2.3%	2.1%	2.5%	**2.8%**	2.0%	3.6%	9.2%	3.7%	5.9%	9.2%	**7.0%**	6.3%	7.6%
Pear	8.7%	6.0%	6.9%	6.6%	**7.0%**	5.9%	8.2%	9.9%	10.2%	8.1%	8.3%	**9.2%**	8.5%	9.8%	2.9%	4.4%	3.1%	3.8%	**3.5%**	2.6%	4.4%	7.0%	3.8%	5.9%	4.7%	**5.4%**	4.8%	5.9%
Orange or mandarin	32.8%	34.7%	34.7%	30.0%	**33.1%**	31.0%	35.1%	33.4%	33.9%	30.3%	31.0%	**32.2%**	31.2%	33.3%	26.8%	24.4%	24.5%	22.6%	**24.7%**	22.6%	26.8%	35.7%	30.6%	29.7%	28.7%	**31.3%**	30.1%	32.4%
Kiwifruit	7.6%	8.3%	5.9%	5.6%	**6.9%**	5.8%	8.0%	10.0%	11.0%	9.6%	10.5%	**10.3%**	9.6%	10.9%	7.8%	8.0%	9.2%	7.2%	**8.1%**	6.8%	9.5%	9.5%	8.1%	9.1%	11.4%	**9.5%**	8.7%	10.2%
Grapes	7.0%	7.0%	7.3%	6.6%	**7.0%**	5.9%	8.1%	10.0%	9.3%	10.0%	8.8%	**9.5%**	8.9%	10.2%	6.7%	8.3%	7.1%	6.3%	**7.1%**	5.9%	8.4%	12.2%	5.6%	4.3%	8.0%	**7.6%**	6.9%	8.3%
Berries	7.4%	6.0%	6.1%	6.6%	**6.5%**	5.4%	7.6%	8.7%	9.7%	8.6%	8.0%	**8.8%**	8.2%	9.4%	13.4%	10.8%	13.4%	13.5%	**12.7%**	11.1%	14.4%	14.3%	11.5%	12.1%	10.0%	**12.1%**	11.2%	12.9%
Banana	21.0%	18.2%	23.6%	21.6%	**21.1%**	19.3%	22.9%	28.9%	26.9%	26.2%	25.7%	**27.0%**	26.0%	28.0%	22.8%	26.2%	21.6%	26.1%	**24.1%**	22.0%	26.1%	24.2%	23.9%	24.6%	28.5%	**25.2%**	24.1%	26.3%
Apple	45.2%	44.0%	42.6%	41.6%	**43.3%**	41.1%	45.5%	44.0%	43.6%	38.6%	38.0%	**41.2%**	40.1%	42.3%	49.0%	43.0%	42.8%	50.6%	**46.1%**	43.7%	48.5%	38.5%	39.5%	35.4%	33.9%	**37.0%**	35.7%	38.2%

Table 10: Numbers of learner's lunches where different food items were <u>available</u> (weighted by learner's age) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

Available (weighted) Food Item	Week 1 (September)														Week 2 (November)												
	Control							Treatment							Control							Treatment						
	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper
Tomato	5.6%	6.1%	5.5%	6.9%	**6.0%**	3.6%	8.5%	7.4%	6.6%	5.6%	5.3%	**6.3%**	4.9%	7.7%	5.1%	7.5%	5.5%	10.6%	**6.9%**	4.5%	9.4%	18.2%	18.6%	17.0%	26.0%	**19.8%**	15.8%	23.8%
Pumpkin	0.8%	0.1%	0.9%	1.3%	**0.8%**	0.2%	1.4%	2.0%	1.3%	1.1%	0.9%	**1.3%**	0.7%	2.0%	0.6%	0.6%	0.2%	1.3%	**0.6%**	0.3%	1.0%	0.5%	0.7%	1.6%	0.9%	**0.9%**	0.4%	1.4%
Potato	3.3%	2.5%	3.4%	3.4%	**3.1%**	2.0%	4.3%	5.3%	4.1%	2.8%	2.5%	**3.8%**	2.6%	4.9%	4.0%	3.1%	3.1%	6.7%	**4.0%**	2.3%	5.8%	19.0%	4.8%	11.4%	8.9%	**11.1%**	8.0%	14.3%
Peas	1.3%	2.2%	1.4%	2.3%	**1.8%**	0.9%	2.6%	2.3%	2.2%	1.4%	1.1%	**1.8%**	1.2%	2.4%	1.4%	1.4%	1.5%	2.6%	**1.7%**	0.8%	2.6%	12.8%	6.7%	11.9%	2.7%	**8.7%**	5.7%	11.6%
Lettuce	9.9%	10.8%	11.6%	10.9%	**10.8%**	7.2%	14.5%	7.8%	5.8%	6.7%	6.3%	**6.7%**	5.1%	8.2%	8.7%	9.6%	10.6%	10.0%	**9.7%**	6.5%	12.9%	28.6%	26.2%	26.0%	34.2%	**28.6%**	23.9%	33.4%
Kumara	0.9%	0.9%	0.8%	1.7%	**1.1%**	0.3%	1.8%	1.9%	2.0%	1.0%	0.9%	**1.5%**	0.9%	2.0%	1.1%	0.6%	0.6%	1.9%	**1.0%**	0.4%	1.6%	0.9%	1.6%	2.7%	0.9%	**1.5%**	0.8%	2.3%
Green leafy veg	1.1%	0.7%	1.6%	1.3%	**1.2%**	0.4%	1.9%	1.8%	1.0%	1.3%	1.0%	**1.3%**	0.9%	1.7%	1.7%	1.1%	0.7%	1.3%	**1.2%**	0.2%	2.2%	9.7%	4.8%	4.1%	8.2%	**6.7%**	5.0%	8.4%
Cucumber	4.2%	6.2%	6.8%	5.1%	**5.6%**	4.1%	7.1%	6.9%	6.6%	7.5%	6.9%	**7.0%**	5.3%	8.7%	7.4%	7.2%	7.7%	6.5%	**7.3%**	4.7%	9.8%	11.5%	18.9%	18.3%	23.5%	**17.9%**	13.7%	22.0%
Corn	1.9%	2.5%	1.6%	2.8%	**2.2%**	1.2%	3.2%	4.0%	3.3%	2.8%	2.9%	**3.3%**	2.3%	4.3%	1.2%	1.9%	1.7%	3.9%	**2.0%**	1.1%	3.0%	19.0%	11.8%	16.9%	4.2%	**13.2%**	9.9%	16.5%
Cauliflower	0.3%	0.3%	0.2%	0.0%	**0.2%**	-0.1%	0.4%	1.0%	0.8%	0.8%	1.2%	**0.9%**	0.4%	1.4%	0.4%	0.6%	0.4%	0.8%	**0.5%**	0.1%	0.9%	1.8%	6.5%	0.8%	0.4%	**2.5%**	1.0%	3.9%
Carrot	16.8%	15.0%	14.0%	11.4%	**14.3%**	10.8%	17.9%	13.9%	13.5%	12.6%	12.1%	**13.1%**	10.6%	15.5%	13.4%	12.1%	13.5%	15.0%	**13.4%**	9.9%	16.8%	50.5%	36.3%	37.9%	38.9%	**41.1%**	36.3%	45.8%
Capsicum	1.5%	2.1%	2.5%	1.8%	**2.0%**	0.8%	3.2%	2.8%	1.9%	1.9%	2.0%	**2.2%**	1.5%	2.8%	1.4%	1.9%	1.9%	2.8%	**2.0%**	0.8%	3.1%	8.5%	11.1%	5.0%	7.5%	**8.1%**	5.5%	10.7%
Cabbage	3.1%	1.8%	2.4%	2.5%	**2.5%**	1.4%	3.5%	2.8%	2.4%	2.0%	1.7%	**2.3%**	1.6%	2.9%	3.4%	1.9%	2.2%	2.1%	**2.4%**	1.3%	3.6%	10.2%	14.3%	9.5%	13.1%	**11.8%**	8.9%	14.6%
Brocolli	1.2%	1.2%	0.7%	1.6%	**1.2%**	0.6%	1.8%	2.6%	1.7%	2.0%	1.6%	**2.0%**	1.3%	2.8%	1.1%	1.2%	0.2%	1.8%	**1.0%**	0.5%	1.6%	3.6%	7.8%	1.6%	1.2%	**3.7%**	2.1%	5.2%
Beans	0.6%	0.5%	0.6%	0.5%	**0.6%**	0.2%	0.9%	1.8%	1.4%	1.1%	1.0%	**1.3%**	0.9%	1.8%	1.1%	0.4%	0.6%	1.0%	**0.8%**	0.1%	1.4%	1.7%	1.1%	2.9%	1.8%	**1.9%**	1.3%	2.5%
Sugar coated snacks	4.3%	3.4%	4.6%	7.5%	**4.9%**	3.2%	6.6%	7.8%	6.8%	6.5%	6.5%	**6.9%**	5.4%	8.5%	4.7%	4.5%	6.3%	3.3%	**4.8%**	2.8%	6.8%	3.0%	2.9%	2.5%	2.0%	**2.6%**	1.9%	3.4%
Popcorn	9.1%	7.5%	9.5%	10.5%	**9.1%**	6.2%	12.0%	11.6%	13.9%	12.7%	11.4%	**12.4%**	10.5%	14.4%	10.5%	11.0%	14.3%	12.0%	**11.9%**	8.7%	15.2%	14.1%	6.9%	10.3%	11.1%	**10.6%**	8.0%	13.2%
Muesli/fruit bar	35.9%	39.2%	37.7%	38.3%	**37.8%**	32.6%	43.0%	38.2%	37.3%	37.0%	37.3%	**37.5%**	35.2%	39.8%	40.8%	40.4%	42.3%	42.8%	**41.5%**	36.9%	46.0%	18.3%	18.6%	17.1%	15.7%	**17.5%**	14.8%	20.2%
Lollies	15.1%	15.1%	19.8%	16.8%	**16.7%**	12.2%	21.3%	19.2%	20.1%	17.7%	18.1%	**18.8%**	16.6%	21.0%	16.8%	13.4%	14.9%	10.5%	**14.2%**	11.3%	17.1%	9.6%	9.6%	8.2%	8.1%	**8.9%**	6.9%	10.9%
Cracker	22.2%	23.9%	23.2%	18.7%	**22.0%**	19.1%	25.0%	26.8%	26.4%	25.2%	24.2%	**25.7%**	22.7%	28.7%	25.3%	25.0%	21.7%	20.7%	**23.4%**	19.7%	27.1%	17.2%	19.5%	12.9%	13.4%	**15.9%**	13.2%	18.5%
Cornchips	11.9%	15.5%	17.8%	18.4%	**15.9%**	13.0%	18.8%	12.8%	11.1%	11.4%	12.4%	**11.9%**	10.0%	13.9%	17.8%	16.3%	13.0%	17.4%	**16.1%**	13.3%	18.8%	14.7%	10.7%	14.4%	14.2%	**13.5%**	11.0%	15.9%
Cookies	41.1%	42.4%	44.6%	42.2%	**42.6%**	39.1%	46.1%	43.9%	43.4%	41.2%	42.5%	**42.8%**	40.0%	45.6%	41.0%	38.1%	38.8%	39.7%	**39.4%**	36.1%	42.7%	23.9%	20.1%	22.5%	19.1%	**21.5%**	18.0%	25.0%
Chocolate	10.6%	8.9%	9.4%	13.7%	**10.6%**	7.5%	13.7%	12.5%	11.6%	10.7%	10.0%	**11.3%**	9.5%	13.0%	9.6%	10.5%	9.8%	10.2%	**10.0%**	6.5%	13.5%	7.4%	7.0%	6.1%	5.4%	**6.5%**	5.0%	8.0%
Potato chips/chippies	55.3%	54.1%	58.8%	59.4%	**56.9%**	52.7%	61.0%	64.6%	63.9%	62.1%	62.2%	**63.3%**	60.6%	66.0%	56.8%	57.0%	54.2%	56.6%	**56.1%**	50.0%	62.2%	35.4%	31.7%	30.8%	32.0%	**32.5%**	28.7%	36.4%
Cake of muffin	14.1%	15.5%	14.5%	12.8%	**14.2%**	11.5%	16.9%	15.5%	13.6%	14.5%	14.0%	**14.4%**	12.9%	15.9%	13.3%	11.9%	14.6%	11.1%	**12.8%**	9.6%	16.1%	28.3%	25.1%	22.0%	30.3%	**26.4%**	21.4%	31.4%
Watermelon	3.7%	2.7%	2.2%	4.1%	**3.2%**	1.9%	4.5%	5.2%	4.6%	3.9%	3.6%	**4.4%**	3.0%	5.8%	2.6%	2.6%	2.9%	1.9%	**2.5%**	1.4%	3.7%	4.2%	1.7%	1.8%	2.6%	**2.6%**	1.7%	3.5%
Stonefruit (peach/plums)	2.0%	4.9%	1.8%	3.1%	**3.0%**	2.0%	3.9%	3.8%	3.6%	3.1%	2.9%	**3.3%**	2.3%	4.4%	5.0%	1.5%	3.2%	3.1%	**3.2%**	1.6%	4.8%	5.3%	4.5%	5.7%	9.8%	**6.2%**	4.0%	8.4%
Pineapple	3.0%	2.7%	2.5%	2.7%	**2.7%**	1.3%	4.1%	3.9%	3.8%	3.9%	3.0%	**3.7%**	2.5%	4.8%	4.0%	2.1%	2.0%	2.1%	**2.6%**	1.5%	3.7%	7.6%	3.4%	7.4%	10.4%	**7.1%**	4.8%	9.4%
Pear	8.2%	5.4%	6.3%	5.9%	**6.4%**	3.9%	9.0%	10.0%	10.3%	8.5%	8.3%	**9.4%**	6.7%	12.0%	2.8%	3.8%	2.8%	3.8%	**3.2%**	1.6%	4.9%	6.0%	3.7%	5.4%	4.6%	**4.9%**	3.3%	6.6%
Orange or mandarin	32.7%	35.1%	33.9%	30.2%	**33.0%**	28.5%	37.6%	33.0%	33.3%	30.3%	30.7%	**31.9%**	29.2%	34.6%	27.2%	25.1%	24.1%	22.1%	**24.9%**	20.1%	29.6%	38.4%	32.1%	29.7%	28.1%	**32.3%**	27.4%	37.1%
Kiwifruit	7.3%	7.8%	5.7%	5.0%	**6.5%**	4.5%	8.5%	9.6%	10.7%	9.4%	10.1%	**9.9%**	7.4%	12.4%	7.7%	7.7%	9.3%	6.9%	**7.9%**	5.0%	10.9%	9.0%	7.5%	8.6%	10.5%	**8.8%**	5.4%	12.3%
Grapes	7.1%	7.3%	7.2%	7.1%	**7.2%**	5.5%	8.9%	10.3%	9.5%	9.9%	9.0%	**9.7%**	7.9%	11.5%	7.0%	8.6%	8.1%	6.7%	**7.7%**	5.8%	9.6%	11.5%	5.5%	4.5%	7.5%	**7.3%**	5.5%	9.1%
Berries	7.5%	5.8%	5.8%	6.1%	**6.3%**	4.9%	7.6%	8.6%	9.9%	8.8%	8.3%	**8.9%**	7.3%	10.6%	12.3%	10.0%	12.8%	12.3%	**11.8%**	9.1%	14.6%	12.4%	10.9%	11.9%	9.9%	**11.3%**	8.6%	14.0%
Banana	20.8%	18.6%	23.7%	21.7%	**21.2%**	18.3%	24.0%	28.3%	25.6%	25.4%	24.8%	**26.1%**	23.2%	29.0%	22.7%	26.2%	22.4%	26.2%	**24.3%**	20.3%	28.3%	23.3%	23.8%	23.7%	27.3%	**24.4%**	20.1%	28.8%
Apple	43.5%	44.2%	42.5%	40.4%	**42.7%**	37.2%	48.2%	43.0%	42.2%	37.5%	36.2%	**39.9%**	36.8%	43.0%	49.1%	42.7%	42.2%	49.9%	**45.8%**	41.3%	50.2%	39.1%	39.9%	35.6%	33.7%	**37.2%**	31.9%	42.6%

Table 11: Numbers of learner's lunches where different food items were consumed (unweighted) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

Consumed (unweighted) Food Item	Week 1 (September)														Week 2 (November)												
	Control							Treatment							Control							Treatment						
	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper
Tomato	4.9%	6.0%	6.1%	7.8%	**6.2%**	5.2%	7.3%	6.2%	6.1%	5.4%	5.4%	**5.8%**	5.3%	6.3%	4.7%	6.4%	5.6%	10.7%	**6.6%**	5.4%	7.8%	16.0%	16.8%	16.9%	22.7%	**18.0%**	17.0%	18.9%
Pumpkin	0.8%	0.2%	1.0%	1.0%	**0.8%**	0.4%	1.1%	1.1%	0.7%	0.7%	0.5%	**0.8%**	0.6%	1.0%	0.7%	0.5%	0.2%	1.6%	**0.7%**	0.3%	1.1%	0.5%	0.5%	1.9%	0.8%	**0.9%**	0.7%	1.2%
Potato	3.3%	2.7%	3.4%	3.5%	**3.2%**	2.4%	4.0%	4.2%	4.0%	2.6%	2.5%	**3.4%**	3.0%	3.8%	3.1%	2.8%	2.8%	6.0%	**3.5%**	2.6%	4.4%	17.3%	4.6%	11.2%	7.9%	**10.3%**	9.6%	11.1%
Peas	1.2%	2.1%	1.4%	1.9%	**1.7%**	1.1%	2.2%	1.8%	1.6%	1.1%	0.8%	**1.4%**	1.1%	1.6%	1.6%	1.4%	1.6%	2.5%	**1.7%**	1.1%	2.4%	10.8%	6.7%	11.8%	2.6%	**8.1%**	7.4%	8.8%
Lettuce	9.5%	10.3%	11.5%	11.1%	**10.6%**	9.2%	11.9%	6.9%	5.2%	6.6%	6.6%	**6.3%**	5.8%	6.9%	9.2%	8.7%	10.6%	10.7%	**9.7%**	8.3%	11.2%	26.6%	24.4%	23.0%	30.9%	**26.1%**	25.0%	27.2%
Kumara	1.2%	0.8%	1.0%	1.6%	**1.2%**	0.7%	1.6%	1.3%	1.5%	0.7%	0.5%	**1.0%**	0.8%	1.2%	1.1%	0.7%	0.7%	1.9%	**1.0%**	0.6%	1.5%	1.0%	1.2%	3.0%	0.6%	**1.5%**	1.2%	1.8%
Green leafy veg	1.4%	0.8%	1.2%	1.6%	**1.3%**	0.8%	1.7%	1.2%	0.7%	1.1%	0.8%	**0.9%**	0.7%	1.2%	2.0%	1.1%	0.9%	0.9%	**1.3%**	0.7%	1.8%	11.2%	5.2%	4.2%	8.6%	**7.3%**	6.7%	8.0%
Cucumber	4.5%	5.8%	7.1%	4.7%	**5.6%**	4.6%	6.6%	5.8%	6.4%	7.1%	6.6%	**6.5%**	5.9%	7.0%	6.9%	6.0%	8.0%	6.3%	**6.8%**	5.6%	8.1%	11.3%	17.9%	16.2%	22.6%	**16.8%**	15.9%	17.7%
Corn	2.1%	2.5%	1.6%	2.9%	**2.3%**	1.6%	2.9%	3.3%	3.0%	2.5%	2.6%	**2.9%**	2.5%	3.2%	1.3%	1.8%	2.1%	4.1%	**2.2%**	1.5%	2.9%	16.8%	10.7%	16.4%	4.2%	**12.2%**	11.4%	13.0%
Cauliflower	0.4%	0.4%	0.2%	0.0%	**0.3%**	0.0%	0.5%	0.6%	0.3%	0.5%	1.0%	**0.6%**	0.4%	0.8%	0.4%	0.5%	0.5%	0.6%	**0.5%**	0.2%	0.8%	1.2%	4.1%	0.5%	0.4%	**1.6%**	1.3%	1.9%
Carrot	15.7%	13.4%	13.7%	11.1%	**13.5%**	12.0%	15.0%	12.2%	12.3%	12.0%	11.9%	**12.1%**	11.4%	12.8%	13.0%	11.5%	12.0%	13.8%	**12.5%**	10.9%	14.1%	45.1%	35.1%	31.5%	35.3%	**37.0%**	35.8%	38.2%
Capsicum	1.4%	1.9%	2.6%	1.4%	**1.9%**	1.3%	2.5%	2.0%	1.4%	1.5%	1.8%	**1.7%**	1.4%	2.0%	1.1%	1.4%	2.1%	2.5%	**1.7%**	1.1%	2.4%	7.6%	9.5%	4.4%	8.7%	**7.6%**	6.9%	8.2%
Cabbage	2.7%	1.7%	3.0%	2.5%	**2.5%**	1.8%	3.1%	2.3%	2.0%	1.7%	1.4%	**1.9%**	1.6%	2.2%	3.4%	2.1%	1.9%	2.5%	**2.5%**	1.7%	3.2%	8.0%	11.7%	7.4%	10.7%	**9.4%**	8.7%	10.2%
Brocolli	1.0%	1.4%	0.8%	0.8%	**1.0%**	0.6%	1.4%	2.0%	1.3%	1.5%	1.2%	**1.5%**	1.2%	1.8%	1.3%	0.9%	0.2%	1.9%	**1.0%**	0.6%	1.5%	3.3%	6.6%	1.3%	1.2%	**3.2%**	2.8%	3.6%
Beans	0.8%	0.4%	0.8%	0.6%	**0.7%**	0.3%	1.0%	1.4%	1.1%	0.9%	0.8%	**1.1%**	0.8%	1.3%	0.9%	0.2%	0.7%	1.3%	**0.7%**	0.3%	1.2%	1.6%	0.8%	3.0%	2.0%	**1.8%**	1.5%	2.1%
Sugar coated snacks	3.9%	3.1%	4.8%	7.0%	**4.7%**	3.7%	5.6%	5.9%	5.1%	5.0%	5.6%	**5.4%**	4.9%	5.9%	4.0%	3.4%	5.9%	3.1%	**4.2%**	3.2%	5.2%	2.7%	2.7%	2.4%	2.1%	**2.5%**	2.1%	2.9%
Popcorn	8.0%	7.0%	9.7%	9.9%	**8.6%**	7.4%	9.9%	10.2%	11.3%	10.8%	10.1%	**10.6%**	9.9%	11.3%	9.8%	9.4%	12.5%	11.3%	**10.7%**	9.2%	12.2%	12.0%	6.7%	9.7%	9.9%	**9.5%**	8.8%	10.3%
Muesli/fruit bar	34.2%	35.3%	34.3%	35.6%	**34.8%**	32.7%	36.9%	33.1%	32.2%	32.1%	33.2%	**32.6%**	31.6%	33.7%	38.9%	37.0%	38.8%	39.3%	**38.5%**	36.1%	40.8%	16.5%	18.2%	15.3%	15.7%	**16.5%**	15.5%	17.4%
Lollies	14.0%	13.8%	19.0%	17.3%	**16.0%**	14.4%	17.6%	16.9%	17.0%	15.5%	15.7%	**16.3%**	15.5%	17.1%	15.9%	13.6%	16.5%	11.3%	**14.5%**	12.8%	16.2%	8.9%	9.3%	8.0%	7.6%	**8.5%**	7.8%	9.2%
Cracker	21.6%	22.1%	22.0%	17.7%	**20.9%**	19.1%	22.7%	21.5%	20.6%	20.2%	20.2%	**20.6%**	19.7%	21.5%	23.7%	22.5%	20.5%	21.1%	**22.0%**	20.0%	24.0%	14.8%	17.4%	11.4%	13.4%	**14.4%**	13.5%	15.2%
Cornchips	11.1%	14.5%	17.4%	18.1%	**15.3%**	13.7%	16.9%	11.8%	10.2%	10.3%	11.8%	**11.0%**	10.3%	11.7%	16.3%	14.0%	12.5%	15.4%	**14.5%**	12.8%	16.2%	12.6%	10.5%	15.1%	13.8%	**12.9%**	12.1%	13.8%
Cookies	39.2%	40.5%	43.6%	41.2%	**41.1%**	39.0%	43.3%	38.9%	38.7%	37.0%	38.0%	**38.2%**	37.1%	39.3%	39.1%	37.0%	38.4%	38.1%	**38.2%**	35.8%	40.5%	23.2%	19.2%	21.9%	17.8%	**20.6%**	19.6%	21.6%
Chocolate	10.1%	7.8%	9.5%	13.8%	**10.2%**	8.9%	11.6%	10.6%	10.2%	9.5%	8.8%	**9.8%**	9.2%	10.5%	9.2%	9.7%	9.4%	9.4%	**9.4%**	8.0%	10.8%	7.2%	7.2%	5.8%	5.7%	**6.5%**	5.9%	7.1%
Potato chips/chippies	53.0%	51.9%	55.8%	57.8%	**54.6%**	52.4%	56.8%	60.1%	59.1%	58.0%	58.4%	**58.9%**	57.8%	60.0%	55.5%	55.9%	53.4%	55.3%	**55.0%**	52.6%	57.4%	33.9%	31.1%	30.5%	31.0%	**31.7%**	30.5%	32.8%
Cake of muffin	13.4%	14.5%	14.7%	12.6%	**13.8%**	12.3%	15.3%	13.6%	11.9%	13.7%	12.8%	**13.0%**	12.2%	13.7%	13.2%	12.0%	14.8%	11.6%	**13.0%**	11.3%	14.6%	23.2%	21.1%	21.6%	26.9%	**23.1%**	22.1%	24.2%
Watermelon	3.9%	2.3%	2.2%	3.7%	**3.0%**	2.3%	3.8%	4.3%	3.4%	3.1%	3.2%	**3.5%**	3.1%	3.9%	2.7%	2.8%	3.1%	1.9%	**2.6%**	1.9%	3.4%	4.4%	1.6%	1.5%	2.7%	**2.6%**	2.2%	3.0%
Stonefruit (peach/plums)	1.9%	3.9%	1.6%	2.5%	**2.5%**	1.8%	3.1%	3.0%	2.8%	2.4%	2.5%	**2.7%**	2.3%	3.1%	4.5%	1.6%	3.3%	3.1%	**3.1%**	2.3%	4.0%	5.8%	3.9%	4.5%	9.5%	**5.9%**	5.3%	6.5%
Pineapple	2.9%	2.1%	2.4%	2.3%	**2.4%**	1.7%	3.1%	3.2%	3.0%	3.7%	2.8%	**3.2%**	2.8%	3.6%	4.0%	2.1%	2.1%	2.5%	**2.7%**	1.9%	3.5%	8.3%	3.5%	5.3%	9.1%	**6.5%**	5.9%	7.1%
Pear	7.4%	5.0%	5.9%	6.2%	**6.1%**	5.1%	7.2%	8.6%	8.4%	7.2%	7.3%	**7.9%**	7.3%	8.5%	2.2%	3.9%	3.1%	3.8%	**3.2%**	2.3%	4.1%	6.6%	3.6%	5.6%	4.5%	**5.1%**	4.6%	5.7%
Orange or mandarin	31.1%	32.9%	32.7%	27.4%	**31.1%**	29.0%	33.1%	28.8%	28.9%	26.7%	27.1%	**27.9%**	26.9%	28.9%	23.9%	22.5%	23.3%	21.1%	**22.8%**	20.8%	24.9%	30.9%	26.9%	25.8%	25.3%	**27.3%**	26.2%	28.5%
Kiwifruit	6.2%	7.6%	5.7%	5.1%	**6.2%**	5.1%	7.2%	8.8%	8.8%	7.6%	9.0%	**8.5%**	7.9%	9.2%	7.2%	7.4%	8.7%	7.2%	**7.6%**	6.3%	8.9%	7.5%	6.8%	6.2%	9.3%	**7.4%**	6.8%	8.1%
Grapes	6.6%	7.0%	7.1%	6.2%	**6.7%**	5.6%	7.8%	9.1%	8.3%	9.2%	8.1%	**8.7%**	8.0%	9.3%	6.5%	7.8%	6.6%	6.3%	**6.8%**	5.6%	8.1%	11.7%	5.3%	4.3%	7.8%	**7.3%**	6.7%	8.0%
Berries	7.0%	5.6%	6.1%	6.4%	**6.3%**	5.2%	7.3%	7.7%	8.8%	7.6%	7.5%	**7.9%**	7.3%	8.5%	13.2%	10.3%	13.2%	13.5%	**12.5%**	10.9%	14.1%	13.5%	11.0%	11.4%	9.5%	**11.4%**	10.6%	12.2%
Banana	17.9%	16.7%	20.4%	19.1%	**18.5%**	16.8%	20.2%	25.5%	23.2%	22.5%	22.6%	**23.5%**	22.6%	24.5%	21.0%	24.6%	19.8%	25.2%	**22.5%**	20.4%	24.5%	20.6%	20.9%	20.6%	24.8%	**21.6%**	20.6%	22.7%
Apple	42.3%	38.8%	39.8%	38.7%	**39.9%**	37.7%	42.0%	37.6%	37.4%	33.5%	33.3%	**35.6%**	34.5%	36.7%	42.5%	38.6%	39.8%	45.3%	**41.3%**	38.9%	43.7%	33.5%	35.3%	31.5%	29.3%	**32.5%**	31.4%	33.7%

Table 12: Numbers of learners lunches where different food items were consumed (weighted) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

Consumed (weighted) Food Item	Week 1 (September)														Week 2 (November)												
	Control							Treatment							Control							Treatment						
	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper	Day 2	Day 3	Day 4	Day 5	Overall	Lower	Upper
Tomato	4.7%	6.1%	5.4%	6.8%	**5.7%**	3.3%	8.2%	6.2%	5.5%	4.8%	4.7%	**5.3%**	4.1%	6.6%	4.9%	7.0%	5.5%	10.6%	**6.7%**	4.3%	9.2%	15.3%	16.8%	15.1%	20.7%	**16.9%**	13.2%	20.6%
Pumpkin	0.8%	0.1%	0.9%	1.3%	**0.8%**	0.2%	1.4%	1.4%	0.8%	0.8%	0.6%	**0.9%**	0.4%	1.5%	0.6%	0.6%	0.2%	1.3%	**0.6%**	0.3%	1.0%	0.4%	0.5%	1.5%	0.9%	**0.8%**	0.3%	1.3%
Potato	3.1%	2.3%	3.4%	3.4%	**3.0%**	1.9%	4.1%	4.0%	3.5%	2.5%	2.3%	**3.1%**	2.1%	4.1%	3.5%	2.9%	3.1%	6.7%	**3.9%**	2.1%	5.6%	16.8%	4.4%	10.5%	8.4%	**10.1%**	7.2%	12.9%
Peas	1.2%	2.0%	1.4%	2.0%	**1.6%**	0.8%	2.5%	1.7%	1.7%	1.1%	0.9%	**1.4%**	0.9%	1.9%	1.4%	1.1%	1.5%	2.6%	**1.6%**	0.7%	2.5%	11.1%	6.2%	11.0%	2.6%	**7.9%**	5.2%	10.5%
Lettuce	9.6%	10.5%	11.6%	10.8%	**10.6%**	7.0%	14.3%	6.7%	4.8%	6.0%	5.8%	**5.8%**	4.3%	7.3%	8.7%	9.3%	10.6%	10.0%	**9.6%**	6.4%	12.8%	25.9%	24.1%	23.1%	30.2%	**25.8%**	21.4%	30.1%
Kumara	0.9%	0.9%	0.8%	1.7%	**1.1%**	0.3%	1.8%	1.4%	1.6%	0.7%	0.6%	**1.1%**	0.7%	1.5%	1.1%	0.6%	0.6%	1.9%	**1.0%**	0.4%	1.6%	0.9%	1.3%	2.5%	0.7%	**1.4%**	0.7%	2.0%
Green leafy veg	1.1%	0.7%	1.3%	1.3%	**1.1%**	0.4%	1.8%	1.3%	0.6%	1.0%	0.7%	**0.9%**	0.6%	1.2%	1.7%	1.1%	0.7%	1.3%	**1.2%**	0.2%	2.2%	9.0%	4.4%	3.5%	8.0%	**6.2%**	4.6%	7.9%
Cucumber	4.0%	6.2%	6.7%	4.5%	**5.4%**	3.9%	6.8%	6.2%	6.1%	7.0%	6.3%	**6.4%**	4.7%	8.0%	6.9%	6.5%	7.5%	6.0%	**6.8%**	4.7%	8.9%	10.9%	17.8%	15.6%	20.6%	**16.1%**	12.1%	20.1%
Corn	1.9%	2.5%	1.6%	2.7%	**2.2%**	1.2%	3.2%	3.2%	2.7%	2.4%	2.6%	**2.7%**	1.8%	3.7%	1.2%	1.9%	1.7%	3.9%	**2.0%**	1.1%	3.0%	17.0%	10.9%	16.2%	3.7%	**12.1%**	9.2%	15.1%
Cauliflower	0.3%	0.3%	0.2%	0.0%	**0.2%**	-0.1%	0.4%	0.6%	0.4%	0.5%	1.0%	**0.6%**	0.2%	1.0%	0.4%	0.4%	0.4%	0.8%	**0.5%**	0.0%	0.9%	1.3%	4.2%	0.5%	0.3%	**1.6%**	0.8%	2.5%
Carrot	15.9%	14.0%	13.5%	10.5%	**13.5%**	10.1%	16.9%	12.2%	12.0%	11.3%	11.3%	**11.7%**	9.3%	14.1%	12.8%	11.5%	12.7%	14.3%	**12.7%**	9.5%	16.0%	46.1%	34.2%	33.7%	35.4%	**37.5%**	33.1%	42.0%
Capsicum	1.5%	2.1%	2.5%	1.8%	**2.0%**	0.8%	3.2%	2.1%	1.4%	1.5%	1.7%	**1.7%**	1.1%	2.3%	1.4%	1.8%	1.9%	2.8%	**1.9%**	0.7%	3.1%	7.4%	8.7%	4.1%	7.2%	**6.9%**	4.7%	9.1%
Cabbage	2.4%	1.7%	2.4%	2.5%	**2.3%**	1.3%	3.2%	2.2%	1.9%	1.6%	1.3%	**1.8%**	1.2%	2.3%	3.2%	1.9%	2.2%	2.1%	**2.4%**	1.3%	3.4%	8.9%	11.6%	8.1%	11.9%	**10.1%**	7.5%	12.8%
Brocolli	1.0%	1.2%	0.6%	0.8%	**0.9%**	0.4%	1.4%	2.1%	1.4%	1.8%	1.3%	**1.6%**	1.0%	2.3%	1.1%	1.0%	0.2%	1.8%	**1.0%**	0.5%	1.5%	3.3%	6.6%	1.3%	1.2%	**3.2%**	1.8%	4.5%
Beans	0.6%	0.5%	0.6%	0.5%	**0.6%**	0.2%	0.9%	1.3%	1.0%	0.8%	0.8%	**1.0%**	0.6%	1.4%	1.1%	0.3%	0.6%	1.0%	**0.7%**	0.1%	1.3%	1.4%	0.8%	2.8%	1.8%	**1.7%**	1.1%	2.3%
Sugar coated snacks	4.1%	3.1%	4.6%	7.3%	**4.7%**	3.1%	6.3%	6.7%	5.7%	5.8%	5.9%	**6.0%**	4.6%	7.5%	3.6%	4.0%	6.1%	3.3%	**4.3%**	2.5%	6.1%	2.7%	2.9%	2.3%	1.9%	**2.5%**	1.7%	3.2%
Popcorn	8.1%	7.0%	9.5%	10.0%	**8.6%**	6.0%	11.1%	10.4%	11.7%	11.5%	10.0%	**10.9%**	9.0%	12.8%	9.9%	9.7%	13.9%	11.8%	**11.3%**	8.1%	14.4%	13.2%	6.5%	10.0%	10.8%	**10.1%**	7.6%	12.7%
Muesli/fruit bar	32.7%	35.8%	35.0%	34.9%	**34.6%**	29.2%	40.0%	33.2%	32.7%	32.3%	32.8%	**32.8%**	30.3%	35.2%	38.3%	36.9%	39.5%	38.4%	**38.3%**	33.3%	43.3%	16.7%	17.4%	16.0%	15.1%	**16.3%**	13.7%	19.0%
Lollies	13.5%	13.5%	18.3%	16.6%	**15.5%**	11.1%	19.8%	16.8%	18.1%	16.1%	16.3%	**16.8%**	15.0%	18.7%	16.0%	13.0%	14.9%	10.3%	**13.8%**	10.8%	16.8%	9.0%	9.5%	7.9%	7.9%	**8.6%**	6.7%	10.5%
Cracker	20.7%	21.9%	21.7%	18.0%	**20.6%**	17.7%	23.5%	23.1%	22.2%	21.2%	21.3%	**22.0%**	19.4%	24.6%	23.6%	23.2%	21.1%	19.7%	**22.1%**	18.2%	25.9%	16.4%	18.4%	12.1%	12.9%	**15.1%**	12.4%	17.7%
Cornchips	10.8%	14.4%	17.1%	17.7%	**15.0%**	12.1%	18.0%	11.8%	10.1%	10.3%	11.7%	**11.0%**	9.0%	12.9%	16.6%	15.5%	12.1%	15.9%	**15.0%**	12.4%	17.7%	13.6%	10.2%	13.4%	13.7%	**12.7%**	10.3%	15.1%
Cookies	38.5%	39.7%	42.2%	40.1%	**40.1%**	36.7%	43.6%	39.6%	39.0%	37.5%	38.4%	**38.6%**	36.3%	41.0%	38.9%	35.5%	37.2%	37.9%	**37.3%**	33.9%	40.8%	22.8%	19.0%	21.5%	17.9%	**20.4%**	17.1%	23.7%
Chocolate	9.6%	7.6%	9.1%	13.2%	**9.8%**	6.6%	13.0%	10.8%	10.1%	9.6%	8.7%	**9.9%**	8.2%	11.5%	9.2%	9.9%	9.1%	10.2%	**9.5%**	6.1%	13.0%	6.7%	6.5%	5.6%	5.1%	**6.0%**	4.5%	7.5%
Potato chips/chippies	50.9%	50.9%	53.0%	56.6%	**52.8%**	48.3%	57.3%	60.1%	59.0%	58.0%	58.0%	**58.8%**	56.2%	61.5%	54.4%	55.4%	52.0%	53.2%	**53.8%**	47.1%	60.5%	33.6%	30.4%	29.3%	31.0%	**31.1%**	27.3%	34.8%
Cake of muffin	13.7%	15.0%	13.7%	12.3%	**13.7%**	11.1%	16.3%	14.1%	12.0%	13.1%	12.9%	**13.0%**	11.5%	14.5%	12.7%	11.5%	14.4%	11.1%	**12.5%**	9.4%	15.6%	25.4%	23.3%	20.9%	28.7%	**24.5%**	20.0%	29.1%
Watermelon	3.7%	2.4%	2.2%	4.1%	**3.1%**	1.8%	4.4%	4.7%	3.7%	3.4%	3.3%	**3.8%**	2.5%	5.1%	2.6%	2.6%	2.7%	1.9%	**2.5%**	1.4%	3.6%	4.0%	1.6%	1.8%	2.5%	**2.5%**	1.6%	3.3%
Stonefruit (peach/plums)	1.5%	3.9%	1.4%	2.9%	**2.4%**	1.6%	3.3%	3.2%	3.1%	2.6%	2.6%	**2.9%**	1.9%	3.9%	4.4%	1.5%	3.2%	3.1%	**3.0%**	1.5%	4.6%	4.8%	3.9%	3.9%	8.6%	**5.2%**	3.4%	7.0%
Pineapple	3.0%	2.3%	2.5%	2.4%	**2.5%**	1.1%	4.0%	3.4%	3.1%	3.6%	2.6%	**3.2%**	2.1%	4.3%	3.8%	1.9%	2.0%	2.1%	**2.5%**	1.4%	3.6%	6.9%	3.2%	6.6%	10.3%	**6.6%**	4.4%	8.9%
Pear	6.8%	4.4%	5.5%	5.6%	**5.5%**	3.0%	8.1%	8.8%	8.7%	7.8%	7.3%	**8.2%**	5.5%	10.9%	2.1%	3.4%	2.8%	3.8%	**2.9%**	1.4%	4.5%	5.6%	3.5%	4.9%	4.4%	**4.6%**	3.1%	6.2%
Orange or mandarin	31.0%	33.2%	31.7%	27.6%	**30.9%**	26.6%	35.2%	28.3%	28.7%	26.5%	26.5%	**27.5%**	24.8%	30.2%	24.2%	23.5%	23.0%	20.6%	**23.0%**	18.6%	27.4%	32.3%	27.6%	25.5%	25.2%	**27.8%**	24.0%	31.6%
Kiwifruit	5.8%	7.1%	5.5%	4.6%	**5.8%**	3.9%	7.7%	8.5%	8.9%	7.8%	8.8%	**8.5%**	6.3%	10.7%	7.1%	7.1%	8.9%	6.9%	**7.5%**	4.6%	10.5%	6.6%	6.3%	5.5%	8.5%	**6.7%**	4.4%	9.1%
Grapes	6.8%	7.3%	7.1%	6.8%	**7.0%**	5.3%	8.7%	9.3%	8.6%	9.2%	8.2%	**8.8%**	7.0%	10.6%	6.8%	8.0%	7.1%	6.7%	**7.2%**	5.4%	8.9%	10.9%	5.3%	4.5%	7.3%	**7.1%**	5.2%	8.9%
Berries	7.0%	5.2%	5.8%	5.7%	**5.9%**	4.4%	7.4%	7.7%	9.2%	7.7%	7.6%	**8.1%**	6.6%	9.6%	12.1%	9.6%	12.7%	12.3%	**11.6%**	8.9%	14.3%	11.6%	10.6%	11.2%	9.5%	**10.7%**	8.1%	13.3%
Banana	17.5%	16.8%	19.7%	18.4%	**18.1%**	14.9%	21.3%	25.2%	22.5%	21.9%	22.1%	**23.0%**	20.2%	25.7%	21.1%	24.8%	20.1%	25.4%	**22.7%**	18.5%	26.9%	19.6%	20.7%	19.5%	23.7%	**20.8%**	17.3%	24.3%
Apple	40.8%	38.9%	39.1%	37.3%	**39.0%**	33.1%	45.0%	36.8%	36.5%	32.7%	32.0%	**34.6%**	31.4%	37.9%	41.7%	38.4%	39.1%	44.2%	**40.6%**	35.8%	45.4%	34.1%	35.9%	31.8%	29.5%	**33.0%**	28.0%	37.9%

Table 13: Numbers (and %) of learners hungry or satiated after lunch (unweighted) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

	Week 1 (September)										Week 2 (November)									
	Control					Treatment					Control					Treatment				
Day	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper
1	511	76.9%	**23.1%**	19.4%	26.7%	1983	78.4%	**21.6%**	19.8%	23.4%	435	76.6%	**23.4%**	19.5%	27.4%	1608	79.4%	**20.6%**	18.6%	22.6%
2	476	80.3%	**19.7%**	16.2%	23.3%	1962	76.2%	**23.8%**	21.9%	25.6%	440	80.5%	**19.5%**	15.8%	23.3%	1592	79.2%	**20.8%**	18.8%	22.8%
3	498	82.9%	**17.1%**	13.8%	20.4%	1942	77.9%	**22.1%**	20.2%	23.9%	432	82.9%	**17.1%**	13.6%	20.7%	1556	81.9%	**18.1%**	16.2%	20.0%
4	489	80.4%	**19.6%**	16.1%	23.2%	1817	79.4%	**20.6%**	18.8%	22.5%	412	85.9%	**14.1%**	10.7%	17.4%	1390	81.7%	**18.3%**	16.2%	20.3%
5	467	81.8%	**18.2%**	14.7%	21.7%	1608	80.7%	**19.3%**	17.4%	21.2%	311	87.5%	**12.5%**	8.8%	16.2%	1354	83.2%	**16.8%**	14.8%	18.8%
Total	2441	80.4%	**19.6%**	18.0%	21.2%	9312	78.4%	**21.6%**	20.7%	22.4%	2030	82.3%	**17.7%**	16.0%	19.3%	7500	81.0%	**19.0%**	18.1%	19.9%

Table 14: Numbers (and %) of learners hungry or satiated after lunch (weighted) in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

	Control					Treatment					Control					Treatment				
Day	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper	n	Full	Hungry	Lower	Upper
1	511	77.9%	**22.1%**	16.6%	27.6%	1983	77.9%	**22.1%**	19.1%	25.1%	435	76.6%	**23.4%**	18.8%	28.1%	1608	78.8%	**21.2%**	18.2%	24.1%
2	476	80.2%	**19.8%**	14.6%	25.0%	1962	76.2%	**23.8%**	20.9%	26.7%	440	80.6%	**19.4%**	14.7%	24.1%	1592	78.7%	**21.3%**	18.2%	24.4%
3	498	83.4%	**16.6%**	12.7%	20.5%	1942	77.3%	**22.7%**	19.6%	25.8%	432	83.5%	**16.5%**	11.1%	21.9%	1556	82.1%	**17.9%**	15.1%	20.8%
4	489	80.4%	**19.6%**	14.2%	25.0%	1817	79.0%	**21.0%**	18.3%	23.7%	412	85.6%	**14.4%**	8.4%	20.3%	1390	81.1%	**18.9%**	15.9%	21.9%
5	467	81.6%	**18.4%**	13.1%	23.7%	1608	80.0%	**20.0%**	17.1%	23.0%	311	87.9%	**12.1%**	6.1%	18.2%	1354	82.9%	**17.1%**	14.1%	20.0%
Total	2441	80.7%	**19.3%**	15.6%	23.0%	9312	78.0%	**22.0%**	20.0%	24.1%	2030	82.5%	**17.5%**	13.4%	21.6%	7500	80.6%	**19.4%**	17.2%	21.5%

Table 15: Mean score (and confidence interval) of learners (by age) according to five PEDS QL constructs and Who-5 (unweighted), in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

		Week 1 (September)								Week 2 (November)							
		n	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Overall	n	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Overall
Control	**Emotional Functioning**	**359**	**63.8**	**71.2**	**63.5**	**66.1**	**65.8**		**66.0**	**292**	**67.2**	**64.6**	**67.3**	**66.3**	**66.8**		**66.5**
	Lower limit		58.1	64.7	57.4	61.2	62.1		63.7		61.4	56.8	60.7	61.0	62.4		64.0
	Upper limit		69.5	77.6	69.6	71.1	69.5		68.2		72.9	72.4	73.8	71.5	71.3		69.1
	Physical Functioning	**361**	**77.0**	**78.4**	**75.2**	**76.4**	**77.9**		**77.1**	**295**	**76.8**	**72.1**	**74.1**	**76.4**	**77.7**		**76.1**
	Lower limit		72.0	73.1	70.5	72.5	75.2		75.3		71.9	65.3	68.5	72.7	74.4		74.1
	Upper limit		82.1	83.7	80.0	80.2	80.6		78.9		81.8	79.0	79.6	80.1	81.0		78.1
	School Functioning	**360**	**65.0**	**64.9**	**64.9**	**66.0**	**65.9**		**65.5**	**292**	**66.9**	**70.1**	**67.9**	**63.2**	**67.1**		**66.7**
	Lower limit		58.9	58.9	60.7	62.0	62.7		63.5		61.3	62.8	62.6	58.3	63.2		64.4
	Upper limit		71.2	70.9	69.2	70.0	69.2		67.5		72.5	77.3	73.1	68.1	71.0		69.0
	Social Functioning	**360**	**67.0**	**77.7**	**69.2**	**74.0**	**76.2**		**73.4**	**292**	**70.3**	**73.4**	**73.0**	**76.2**	**75.5**		**74.1**
	Lower limit		60.9	71.8	62.7	69.5	72.9		71.2		64.1	65.3	67.9	71.8	71.9		71.9
	Upper limit		73.1	83.7	75.6	78.5	79.4		75.6		76.4	81.6	78.2	80.5	79.1		76.4
	Total Functioning	**359**	**68.2**	**73.0**	**68.2**	**70.6**	**71.4**		**70.5**	**290**	**70.3**	**70.1**	**71.1**	**70.6**	**71.8**		**71.0**
	Lower limit		63.8	68.3	64.4	67.0	68.9		68.9		65.9	64.1	66.6	67.0	68.7		69.2
	Upper limit		72.7	77.8	72.0	74.2	73.9		72.1		74.8	76.0	75.6	74.2	74.9		72.8
	Who Wellbeing	**364**		**71.5**	**69.3**	**62.1**	**64.3**	**64.4**	**65.6**	**301**		**69.3**	**69.2**	**65.4**	**63.9**	**59.7**	**64.9**
	Lower limit			65.9	64.9	57.2	60.9	59.2	63.5			61.2	63.3	59.7	59.8	53.0	62.3
	Upper limit			77.0	73.8	67.1	67.6	69.6	67.6			77.5	75.1	71.1	68.0	66.4	67.5
Treatment	**Emotional Functioning**	**1361**	**63.4**	**63.0**	**62.8**	**64.6**	**65.0**		**63.6**	**1094**	**65.3**	**66.7**	**69.3**	**70.5**	**65.4**		**67.7**
	Lower limit		60.6	60.7	60.6	61.9	62.2		62.5		62.0	64.2	66.7	67.7	61.7		66.4
	Upper limit		66.1	65.3	65.0	67.3	67.8		64.7		68.6	69.3	71.9	73.3	69.2		69.0
	Physical Functioning	**1377**	**72.2**	**73.1**	**74.7**	**76.5**	**74.6**		**74.2**	**1107**	**74.8**	**75.3**	**78.7**	**79.2**	**75.7**		**76.9**
	Lower limit		69.6	70.8	72.9	74.4	72.4		73.2		72.0	73.0	76.8	76.9	73.0		75.8
	Upper limit		74.7	75.4	76.5	78.5	76.9		75.2		77.5	77.5	80.7	81.5	78.4		77.9
	School Functioning	**1351**	**66.1**	**63.6**	**64.8**	**64.7**	**63.3**		**64.6**	**1092**	**68.7**	**66.8**	**67.1**	**70.3**	**65.4**		**67.8**
	Lower limit		63.5	61.4	62.9	62.3	60.6		63.5		65.7	64.3	64.8	67.7	62.3		66.6
	Upper limit		68.7	65.8	66.8	67.2	66.1		65.6		71.6	69.3	69.5	72.9	68.4		69.0
	Social Functioning	**1360**	**66.1**	**65.6**	**70.0**	**71.3**	**70.6**		**68.6**	**1095**	**68.9**	**69.6**	**73.4**	**75.3**	**71.4**		**71.8**
	Lower limit		63.2	63.2	67.8	68.7	67.6		67.5		65.8	67.0	70.9	72.4	67.9		70.5
	Upper limit		69.0	68.1	72.1	73.9	73.5		69.7		72.0	72.1	75.8	78.1	74.8		73.0
	Total Functioning	**1337**	**66.7**	**66.3**	**68.2**	**69.3**	**68.3**		**67.7**	**1078**	**69.3**	**69.6**	**72.1**	**73.8**	**69.4**		**71.0**
	Lower limit		64.6	64.5	66.6	67.4	66.2		66.9		66.9	67.6	70.2	71.7	66.9		70.0
	Upper limit		68.9	68.1	69.8	71.3	70.4		68.6		71.8	71.6	74.0	75.9	71.9		72.0
	Who Wellbeing	**1193**		**69.0**	**66.0**	**64.4**	**61.7**	**61.1**	**65.4**	**958**		**67.1**	**68.4**	**65.7**	**58.5**	**66.8**	**65.9**
	Lower limit			66.6	64.1	62.1	58.8	56.5	64.3			64.4	66.0	62.8	55.1	62.1	64.5
	Upper limit			71.4	68.0	66.7	64.6	65.7	66.6			69.8	70.8	68.5	61.8	71.6	67.2

Table 16: Mean score (and confidence interval) of learners (by age) according to five PEDS QL constructs and Who-5 (weighted), in September and November for schools in the programme (treatment) and not yet in the programme (control), with robust estimates for "overall".

		Week 1 (September)								Week 2 (November)							
		n	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Overall	n	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Overall
Control	**Emotional Functioning**	**359**	**63.8**	**71.2**	**63.5**	**66.1**	**65.8**		**66.0**	**292**	**67.2**	**64.6**	**67.3**	**66.3**	**66.8**		**66.6**
	Lower limit		56.7	59.3	58.4	60.7	58.9		62.6		58.2	57.4	62.2	61.3	58.1		62.6
	Upper limit		70.9	83.0	68.6	71.6	72.7		69.3		76.1	71.8	72.3	71.3	75.6		70.5
	Physical Functioning	**361**	**77.0**	**78.4**	**75.2**	**76.4**	**77.9**		**77.2**	**295**	**76.8**	**72.1**	**74.1**	**76.4**	**77.7**		**76.2**
	Lower limit		69.9	68.5	70.5	70.9	73.9		74.6		68.6	62.7	66.4	72.6	73.2		73.0
	Upper limit		84.2	88.3	79.9	81.8	81.9		79.8		85.1	81.5	81.7	80.3	82.2		79.4
	School Functioning	**360**	**65.0**	**64.9**	**64.9**	**66.0**	**65.9**		**65.5**	**292**	**66.9**	**70.1**	**67.9**	**63.2**	**67.1**		**66.7**
	Lower limit		58.9	56.2	60.9	60.3	60.5		62.6		60.2	61.0	65.1	57.1	60.4		63.3
	Upper limit		71.2	73.5	69.0	71.6	71.3		68.5		73.6	79.1	70.6	69.3	73.8		70.2
	Social Functioning	**360**	**67.0**	**77.7**	**69.2**	**74.0**	**76.2**		**73.5**	**292**	**70.3**	**73.4**	**73.0**	**76.2**	**75.5**		**74.2**
	Lower limit		62.3	67.6	64.8	68.3	70.6		70.5		62.1	66.3	68.6	71.5	69.1		70.7
	Upper limit		71.7	87.9	73.5	79.6	81.8		76.5		78.4	80.6	77.4	80.8	81.9		77.6
	Total Functioning	**359**	**68.2**	**73.0**	**68.2**	**70.6**	**71.4**		**70.5**	**290**	**70.3**	**70.1**	**71.1**	**70.6**	**71.8**		**71.0**
	Lower limit		62.5	63.5	64.7	65.7	66.4		67.9		62.8	62.5	67.9	66.9	65.7		68.0
	Upper limit		74.0	82.6	71.7	75.5	76.3		73.1		77.9	77.6	74.4	74.3	77.9		74.1
	Who Wellbeing	**364**		**71.5**	**69.3**	**62.1**	**64.3**	**64.4**	**65.2**	**301**		**69.3**	**69.2**	**65.4**	**63.9**	**59.7**	**63.9**
	Lower limit			65.2	64.9	57.3	59.1	56.8	62.0			58.9	64.5	60.5	57.6	49.1	58.7
	Upper limit			77.7	73.7	66.9	69.4	72.0	68.5			79.7	73.9	70.3	70.2	70.3	69.1
Treatment	**Emotional Functioning**	**1361**	**63.4**	**63.0**	**62.8**	**64.6**	**65.0**		**63.6**	**1094**	**65.3**	**66.7**	**69.3**	**70.5**	**65.4**		**67.6**
	Lower limit		59.4	60.2	60.4	61.9	62.1		62.1		60.8	63.4	66.2	67.1	60.6		65.6
	Upper limit		67.3	65.8	65.2	67.3	67.9		65.2		69.9	70.0	72.4	73.8	70.3		69.6
	Physical Functioning	**1377**	**72.2**	**73.1**	**74.7**	**76.5**	**74.6**		**74.2**	**1107**	**74.8**	**75.3**	**78.7**	**79.2**	**75.7**		**76.8**
	Lower limit		68.9	69.9	72.1	74.2	71.7		72.6		70.9	72.6	76.0	76.1	72.3		75.0
	Upper limit		75.5	76.3	77.3	78.7	77.6		75.7		78.6	77.9	81.5	82.2	79.1		78.6
	School Functioning	**1351**	**66.1**	**63.6**	**64.8**	**64.7**	**63.3**		**64.5**	**1092**	**68.7**	**66.8**	**67.1**	**70.3**	**65.4**		**67.8**
	Lower limit		63.3	60.8	62.2	62.1	60.2		63.2		65.2	63.7	64.3	67.3	61.0		66.0
	Upper limit		68.9	66.3	67.4	67.4	66.5		65.9		72.1	69.9	70.0	73.3	69.7		69.5
	Social Functioning	**1360**	**66.1**	**65.6**	**70.0**	**71.3**	**70.6**		**68.6**	**1095**	**68.9**	**69.6**	**73.4**	**75.3**	**71.4**		**71.7**
	Lower limit		61.8	63.0	67.4	68.6	66.6		66.9		64.6	66.3	70.2	72.3	66.8		69.7
	Upper limit		70.4	68.2	72.5	74.0	74.5		70.2		73.3	72.9	76.6	78.2	75.9		73.7
	Total Functioning	**1337**	**66.7**	**66.3**	**68.2**	**69.3**	**68.3**		**67.7**	**1078**	**69.3**	**69.6**	**72.1**	**73.8**	**69.4**		**70.9**
	Lower limit		63.7	64.0	66.1	67.2	65.5		66.4		65.7	66.8	69.5	71.3	65.8		69.3
	Upper limit		69.8	68.6	70.3	71.4	71.1		69.0		73.0	72.3	74.8	76.4	73.1		72.6
	Who Wellbeing	**1193**		**69.0**	**66.0**	**64.4**	**61.7**	**61.1**	**65.0**	**958**		**67.1**	**68.4**	**65.7**	**58.5**	**66.8**	**65.7**
	Lower limit			66.0	64.2	62.0	58.3	55.9	63.3			64.0	65.5	62.8	54.0	63.1	64.0
	Upper limit			72.0	67.9	66.8	65.2	66.3	66.6			70.2	71.3	68.5	63.0	70.5	67.5

Table 17: % of learners designated "at risk" (weighted by age), based on levels determeinted clinically important for a USA adolescent population[30], with robust 95% CI.

Peds-QL	Week 1 (September)										Week 2 (November)									
	Cutoff Score*	Control					Treatment					Control					Treatment				
		n	Above	Below	Lower	Upper	n	Above	Below	Lower	Upper	n	Above	Below	Lower	Upper	n	Above	Below	Lower	Upper
Emotional Functioning	59.57	359	59.6%	**40.4%**	34.6%	46.1%	1361	56.3%	**43.7%**	40.8%	46.7%	292	60.3%	**39.7%**	32.3%	47.1%	1094	56.6%	**43.4%**	39.9%	46.9%
Physical Functioning	72.98	361	58.6%	**41.4%**	36.0%	46.8%	1377	50.3%	**49.7%**	46.1%	53.4%	295	54.8%	**45.2%**	39.7%	50.8%	1107	55.3%	**44.7%**	40.8%	48.6%
School Functioning	62.99	360	57.3%	**42.7%**	35.5%	49.9%	1351	48.5%	**51.5%**	48.3%	54.7%	292	55.7%	**44.3%**	38.1%	50.6%	1092	54.1%	**45.9%**	42.3%	49.5%
Social Functioning	66.6	360	55.8%	**44.2%**	37.9%	50.4%	1360	48.8%	**51.2%**	47.9%	54.6%	292	57.6%	**42.4%**	35.3%	49.6%	1095	50.7%	**49.3%**	45.6%	53.0%
Health Quality of Life	69.71	359	56.8%	**43.2%**	37.4%	49.0%	1337	45.5%	**54.5%**	50.5%	58.5%	290	54.1%	**45.9%**	37.9%	54.0%	1078	52.3%	**47.7%**	43.4%	52.0%

Table 18: Percentage of learner's lunches with at least one food item throughout four days (weighted by age) with robust 95% CI.

	Week 1 (September)										Week 2 (November)									
	Control					Treatment					Control					Treatment				
	n	At least one	None	Lower	Upper	n	At least one	None	Lower	Upper	n	At least one	None	Lower	Upper	n	At least one	None	Lower	Upper
Vegetables	1992	29.1%	**70.9%**	64.8%	77.1%	7616	26.2%	**73.8%**	70.0%	77.5%	1625	30.7%	**69.3%**	62.8%	75.8%	6173	66.8%	**33.2%**	27.2%	39.1%
Snacks	1992	89.0%	**11.0%**	8.0%	14.1%	7616	88.5%	**11.5%**	9.9%	13.0%	1625	90.0%	**10.0%**	7.2%	12.8%	6173	73.8%	**26.2%**	23.0%	29.4%
Fruit	1992	72.3%	**27.7%**	24.0%	31.4%	7616	72.5%	**27.5%**	24.5%	30.4%	1625	74.3%	**25.7%**	20.8%	30.7%	6173	71.1%	**28.9%**	25.3%	32.5%

30

Appendix C: Learner Food Diary Survey (Longitudinal analysis tabled results)

The following tables present the results from the regression analyses, which were undertaken on the longitudinal data derived from the learner food diaries. The longitudinal data represent individual learners' responses from daily surveys over five days in September and again daily surveys over five days in November. The results demonstrate the factors that influenced food consumption, satiety and wellbeing, and these confounders were subsequently controlled for in the full analysis.

As discussed in the methods, a random intercept at the class level is included in addition to fixed effect confounding factors using linear mixed effects modelling. In the following, significance is determined using the Satterwaite method with adjusted degrees of freedom, while confidence intervals are based on bootstrapping (percentile method). Pseudo *R*-squared values are calculated using the method of Nakagawa.[31]

When controlling for confounders, factors were included that are potential risk factors for the outcome and are distributed unequally. In addition, key variables relevant to the theory of the intervention including age and Māori ethnicity were included.

Table 19: Factors (entered individually) influencing change in average number of vegetables consumed between September and November (n=1944)

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	0.98	0.13	7.62	<0.001	103	0.73	1.22
Age	-0.01	0.02	-0.35	0.73	596	-0.04	0.03
Sex (Female vs Male)	0.09	0.04	2.33	0.02	1836	0.02	0.17
Māori	0.06	0.04	1.29	0.20	1880	-0.03	0.14
Pasifika	0.11	0.06	1.95	0.05	1846	0.01	0.23
Pākehā	-0.01	0.04	-0.15	0.88	1884	-0.08	0.09
Background food programmes*							
Targeted (v none)	-0.17	0.42	-0.39	0.70	113	-1.13	0.69
Universal (v none)	-0.29	0.46	-0.64	0.53	112	-1.25	0.58
Food policy in place	0.50	0.13	3.88	<0.001	101	0.26	0.76
Delivery model (Internal v External)	0.59	0.15	3.81	<0.001	104	0.29	0.91
Equity index	0.40	0.13	3.05	0.003	105	0.13	0.67

*Targeted food programmes include KidsCan, and targeted free lunch programmes; Universal food programmes include Free Milk in Schools, Free Fruit in Schools, and Free lunch for all programmes.

[31] Using *MuMIn* R package (author: Kamil Barton) using method of Nakagawa, S., Schielzeth, H. (2013) A general and simple method for obtaining R from Generalized Linear Mixed-effects Models. Methods in Ecology and Evolution 4: 133–142.

Table 20: Net effect of the pilot's influence on consumption of vegetables while controlling for confounders using mixed modelling (n=1944).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-3.06	1.08	-2.82	0.01	116	-5.38	-0.91
Treatment (pilot)	0.88	0.12	7.20	<0.001	103	0.64	1.13
Age	0.02	0.02	0.92	0.36	406	-0.02	0.05
Māori	0.04	0.04	0.81	0.42	1910	-0.06	0.13
Equity index	0.28	0.11	2.62	0.01	108	0.06	0.52
Food policy in place	0.35	0.10	3.35	<0.001	103	0.12	0.56

Random effect: $ICC_{adj.}$(class) = 0.23. $R^2_{marg.}$ = 0.18, $R^2_{cond.}$ = 0.37.

Table 21: Factors (entered individually) influencing change in average number of snacks and sweets consumed between September and November (n = 1944).

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	-0.57	0.11	-5.01	<0.001	101	-0.78	-0.33
Age	0.00	0.02	0.06	0.95	280	-0.04	0.04
Sex (Female v Male)	0.04	0.05	0.89	0.37	1866	-0.05	0.14
Māori	-0.04	0.05	-0.73	0.46	1919	-0.13	0.06
Pasifika	-0.03	0.07	-0.44	0.66	1883	-0.16	0.11
Pākehā	0.04	0.05	0.71	0.48	1919	-0.07	0.13
Background food programmes*							
Targeted (v none)	0.82	0.33	2.51	0.01	106	0.13	1.42
Universal (v none)	0.49	0.35	1.40	0.16	108	-0.31	1.21
Food policy in place	0.10	0.11	0.93	0.36	101	-0.10	0.32
School Model (Internal v External)	0.15	0.13	1.12	0.26	108	-0.11	0.39
Equity Index	0.10	0.11	0.95	0.34	107	-0.10	0.31

*Targeted food programmes include KidsCan, and targeted free lunch programmes; Universal food programmes include Free Milk in Schools, Free Fruit in Schools, and Free lunch for all programmes.

Table 22: Net effect of the pilot's influence on consumption of snacks and sweets while controlling for confounders using mixed modelling (n=1944).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-0.71	0.37	-1.92	0.06	122	-1.41	0.08
Treatment (pilot)	-0.52	0.13	-4.11	<0.001	100	-0.80	-0.26
Age	-0.01	0.02	-0.52	0.60	280	-0.05	0.03
Māori	-0.03	0.05	-0.62	0.53	1938	-0.14	0.08
Background food programmes*							
Targeted (v none)	0.82	0.36	2.30	0.02	99	0.07	1.54
Universal (v none)	0.72	0.31	2.34	0.02	100	0.08	1.33

Random effect $ICC_{adj.}$(class) = 0.13. $R^2_{marg.}$ = 0.05, $R^2_{cond.}$ = 0.17.

Table 23: Factors (entered individually) influencing change in average number of fruits consumed between September and November (n = 1944).

Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	95% C.I.for B lower	upper
Treatment (pilot)	0.02	0.09	0.26	0.80	96	-0.14	0.20
Age	0.03	0.01	1.92	0.06	210	0.00	0.06
Sex (Female v Male)	-0.05	0.04	-1.21	0.23	1882	-0.13	0.02
Māori	-0.06	0.04	-1.27	0.20	1907	-0.15	0.03
Pasifika	0.02	0.06	0.38	0.71	1901	-0.09	0.14
Pākehā	0.04	0.04	1.02	0.31	1893	-0.04	0.14
Background food programmes*							
Targeted (v none)	0.55	0.23	2.40	0.02	93	0.11	0.97
Universal (v none)	0.33	0.25	1.32	0.19	96	-0.17	0.77
Food policy in place	0.09	0.07	1.26	0.21	95	-0.06	0.23
School Model (Internal v Externall)	0.16	0.09	1.75	0.08	102	-0.02	0.33
Equity Index	-0.02	0.08	-0.33	0.74	100	-0.18	0.11

Table 24: Net effect of the pilot's influence on consumption of fruits while controlling for confounder variables using mixed modelling (n=1944).

Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	95% C.I.for B lower	upper
(Intercept)	-0.43	0.18	-2.36	0.02	159	-0.81	-0.04
Treatment (pilot)	0.06	0.09	0.62	0.54	97	-0.12	0.24
Age	0.03	0.01	1.69	0.09	208	0.00	0.05
Māori	-0.07	0.04	-1.64	0.10	1929	-0.15	0.01

Random effect $ICC_{adj.}$(class) = 0.11. $R^2_{marg.}$ = 0.01, $R^2_{cond.}$ = 0.12.

Table 25: Factors (entered individually) influencing change in average satiety between September and November (n=2008).

Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	95% C.I.for B lower	upper
Treatment (pilot)	3.5	1.7	2.1	0.04	99	0.8	6.9
Age	0.5	0.3	1.6	0.11	171	-0.1	1.1
Sex (Female v Male)	1.1	1.0	1.1	0.26	1969	-0.8	3.3
Māori	2.1	1.1	1.9	0.05	1856	-0.1	4.4
Pasifika	3.0	1.5	2.1	0.04	1981	-0.1	6.1
Pākehā	-2.2	1.0	-2.1	0.04	1818	-4.5	-0.2
Background food programmes*							
Targeted (v none)	-8.0	4.5	-1.8	0.08	89	-17.4	1.8
Universal (v none)	-11.0	4.9	-2.2	0.03	95	-21.0	-0.6
Food policy in place	-2.5	4.1	1.2	0.24	97	-9.9	5.9
School Model (Internal v External)	3.7	1.8	2.1	0.04	109	0.3	7.3
Equity index	1.7	1.5	2.8	0.01	103	-1.5	4.6

Table 26: Net effect of the pilot's influence on change in average satiety while controlling for confounder variables using mixed modelling (n=2008).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-47.1	15.6	-3.0	0.003	113	-82.7	-16.1
Treatment (pilot)	3.8	1.8	2.1	0.03	95	0.2	7.0
Age	0.8	0.3	2.4	0.02	185	0.1	1.4
Background food programmes*							
Targeted (v none)	-9.0	4.9	-1.8	0.07	82	-18.4	1.3
Universal (v none)	-11.1	4.3	-2.6	0.01	81	-19.6	-3.1
Māori	1.3	1.1	1.2	0.23	1954	-0.9	3.3
Pākehā	-1.6	1.1	-1.5	0.12	1902	-3.8	0.4
Equity index	0.1	0.0	3.3	0.001	110	0.0	0.2

Random effect $ICC_{adj.}$(class) = 0.04. $R^2_{marg.}$ = 0.02, $R^2_{cond.}$ = 0.06.

Table 27: Factors (entered individually) influencing change in Health Quality of Life between September and November (n=1157*).

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	2.59	1.06	2.44	0.02	70	0.49	4.53
Age	-0.11	0.32	-0.36	0.72	130	-0.77	0.47
Sex (Female v Male)	2.38	0.78	3.07	0.002	1136	0.70	3.91
Māori	0.64	0.82	0.78	0.43	900	-1.06	2.30
Pasifika	2.45	1.18	2.09	0.04	1135	0.17	4.64
Pākehā	1.68	0.80	2.11	0.04	907	0.24	3.31
Background food programmes*							
Targeted (v none)	1.94	2.20	0.88	0.38	43	-2.34	6.76
Universal (v none)	3.53	2.50	1.41	0.16	49	-1.16	8.58
School Model (Internal v External)	-0.47	1.19	-0.40	0.69	59	-2.98	2.06
Equity index	-0.37	0.91	-0.40	0.69	60	-2.26	1.52

*Health Quality of Life questions were limited to those aged 8 to 12 years old, as per the recommended methodology. This applies to all the constructs (physical, emotional, school, social).

Table 28: Net effect of the pilot's influence on measured change in Health Quality of Life while controlling for confounder variables using mixed modelling (n=1157).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-1.71	3.47	-0.49	0.62	134	-8.36	5.71
Treatment (pilot)	2.97	1.08	2.74	0.01	66	0.89	4.96
Sex (Female v Male)	2.28	0.77	2.96	0.00	1150	0.91	3.89
Age	-0.10	0.32	-0.30	0.76	148	-0.77	0.56
Māori	1.22	0.84	1.46	0.14	973	-0.32	2.82
Pākehā	1.95	0.82	2.38	0.02	1039	0.46	3.51

Random effect $ICC_{adj.}$(class) = 0.02. $R^2_{marg.}$ = 0.02, $R^2_{cond.}$ = 0.04.

Table 29: Factors (entered individually) influencing change in Emotional functioning between September and November (n=1157).

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	3.01	1.52	1.98	0.05	60	-0.02	5.98
Age	0.26	0.47	0.56	0.58	122	-0.63	1.29
Sex (Female v Male)	2.43	1.21	2.00	0.05	1136	-0.08	4.81
Māori	1.87	1.26	1.48	0.14	776	-0.27	4.25
Pasifika	5.05	1.82	2.77	0.01	1124	1.25	8.41
Pākehā	-0.32	1.22	-0.26	0.80	762	-2.80	1.94
Background food programmes*	-1.59	3.18	-0.50	0.62	42	-8.08	4.55
Targeted (v none)	1.05	3.64	0.29	0.77	49	-6.22	7.68
Universal (v none)	1.58	1.68	0.94	0.35	52	-1.74	5.11
School Model (Internal v External)	-3.36	3.93	-0.86	0.40	51	-10.76	4.04

Table 30: Net effect of the pilot's influence on measured change in Emotional functioning while controlling for confounder variables using mixed modelling (n=1157).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-5.14	5.05	-1.02	0.31	117	-15.76	4.86
Treatment (pilot)	3.45	1.53	2.26	0.03	60	0.18	6.28
Age	0.36	0.46	0.79	0.04	1151	-0.49	1.25
Sex (Female v Male)	2.54	1.21	2.10	0.43	124	0.28	5.14
Māori	1.93	1.25	1.54	0.12	774	-0.52	4.17

Random effect $ICC_{adj.}$(class) = 0.01. $R^2_{marg.}$ = 0.01, $R^2_{cond.}$ = 0.02.

Table 31: Factors (entered individually) influencing change in Physical functioning between September and November (n=1157).

Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	95% C.I.for B lower	upper
Treatment (pilot)	3.09	1.61	1.92	0.06	78	0.25	6.37
Age	0.10	0.46	0.22	0.83	174	-0.84	0.98
Sex (Female v Male)	2.92	1.03	2.82	0.004	1130	0.80	5.16
Māori	-0.24	1.11	-0.22	0.83	1048	-2.38	1.90
Pasifika	1.56	1.57	1.00	0.32	1133	-1.59	4.69
Pākehā	2.79	1.07	2.61	0.01	1042	0.79	4.93
Background food programmes*							
Targeted (v none)	1.97	3.49	0.56	0.57	57	-4.96	8.99
Universal (v none)	2.86	3.93	0.73	0.47	62	-4.75	11.41
School Model (Internal v External)	-2.49	1.78	-1.40	0.16	73	-6.25	0.89
Equity index	-1.58	1.37	-1.15	0.25	75	-4.13	1.12

Table 32: Net effect of the pilot's influence on measured change in Physical functioning while controlling for confounder variables using mixed modelling (n=1157).

Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	95% C.I.for B lower	upper
(Intercept)	-5.77	4.96	-1.16	0.25	169	-15.72	3.58
Treatment (pilot)	3.69	1.60	2.30	0.02	78	0.66	6.75
Age	0.14	0.45	0.30	0.76	187	-0.69	1.04
Sex (Female v Male)	2.78	1.03	2.71	0.01	1145	0.67	4.58
Māori	0.51	1.13	0.45	0.65	1068	-1.64	2.75
Pākehā	2.77	1.10	2.52	0.01	1109	0.59	5.03

Random effect $ICC_{adj.}$(class) = 0.04. $R^2_{marg.}$ = 0.02, $R^2_{cond.}$ = 0.06.

Table 33: Factors (entered individually) influencing change in Social functioning between September and November (n=1157).*

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df	lower	upper
Treatment (pilot)	2.19	1.39	1.57	0.12	1	-0.54	4.91
Age	-0.34	0.43	-0.80	0.42	1	-1.17	0.49
Sex (Female v Male)	3.66	1.15	3.19	0.00	1	1.41	5.91
Māori	1.80	1.19	1.51	0.13	1	-0.53	4.13
Pasifika	2.31	1.72	1.35	0.18	1	-1.05	5.67
Pākehā	1.56	1.15	1.36	0.17	1	-0.69	3.82
Background food programmes							
Targeted (v none)	-0.69	3.25	-0.21	0.83	1	-7.06	5.68
Universal (v none)	3.16	2.83	1.11	0.27	1	-2.39	8.70
School Model (Internal v External)	1.01	1.53	0.66	0.51	1	-1.98	4.00
Equity index	0.54	1.18	0.46	0.65	1	-1.77	2.85

*Results for GLM model with robust confidence intervals. as data did not support using classroom as a random effect (very low variance explained). Parameters were consistent between the two approaches however.

Table 34: Net effect of the pilot's influence on measured change in Social functioning while controlling for confounder variables (Random effect $ICC_{adj.}$(class) = 0.04. $R^2_{marg.}$ = 0.02, $R^2_{cond.}$ = 0.06.

) using generalised linear model (n=1157).*

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df	lower	upper
(Intercept)	0.87	4.71	0.18	0.85	1	-8.35	10.09
treatment_control1	2.28	1.41	1.62	0.10	1	-0.47	5.04
demo_age_1_1	-0.24	0.43	-0.57	0.57	1	-1.09	0.60
demo_x2_maori1	1.73	1.19	1.46	0.15	1	-0.60	4.05
demo_sex2	3.73	1.15	3.25	0.00	1	1.48	5.98

*Results are reported for GLM model with robust confidence intervals as data did not support using classroom as a random effect (very low variance explained). Parameters were consistent between the two approaches however. Pseudo-R^2 = 0.01.

Table 35: Factors (entered individually) influencing change in School functioning between September and November (n=1157).

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	2.07	1.62	1.28	0.21	78	-0.76	5.49
Age	-0.57	0.47	-1.23	0.22	165	-1.49	0.35
Sex (Female v Male)	-0.34	1.11	-0.31	0.76	1134	-2.57	1.67
Māori	-0.22	1.17	-0.19	0.85	999	-2.71	2.35
Pasifika	1.85	1.67	1.11	0.27	1136	-1.62	5.20
Pākehā	1.54	1.14	1.36	0.17	990	-0.64	3.79
Background food programmes*							
Targeted (v none)	5.35	3.28	1.63	0.11	54	-1.24	11.55
Universal (v none)	8.76	3.72	2.35	0.02	61	1.49	15.50
School Model (Internal v External)	-0.43	1.80	-0.24	0.81	74	-3.74	3.13
Equity index	0.17	1.37	0.13	0.90	74	-2.92	2.71

Table 6: Net effect of the pilot's influence on measured change in School functioning while controlling for confounder variables using mixed modelling (n=1157).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df (adj.)	lower	upper
(Intercept)	-1.51	6.00	-0.25	0.80	101	-12.86	10.64
treatment_control1	2.31	1.65	1.40	0.17	75	-0.77	5.53
demo_age_1_1	-0.47	0.47	-1.00	0.32	166	-1.57	0.52
demo_x2_maori1	-0.05	1.18	-0.05	0.96	1037	-2.51	2.05
Background food programmes*							
Targeted (v none)	6.53	3.47	1.88	0.06	58	0.11	13.65
Universal (v none)	9.34	3.89	2.40	0.02	63	1.91	16.66

Random effect $ICC_{adj.}$(class) = 0.03. $R^2_{marg.}$ = 0.01, $R^2_{cond.}$ = 0.04.

Table 37: Factors (entered individually) influencing change in mental wellbeing (WHO-5) between September and November (n=1305).

						95% C.I.for B	
Influencing Factors (individual)	Beta	SE	Wald	p	df (adj.)	lower	upper
Treatment (pilot)	-1.07	1.39	-0.77	0.44	65	-3.73	1.57
Age	0.40	0.35	1.14	0.26	120	-0.34	1.07
Sex (Female v Male)	-0.29	1.09	-0.27	0.79	1302	-2.41	1.56
Māori	0.10	1.14	0.09	0.93	1020	-2.06	2.63
Pasifika	1.04	1.57	0.66	0.51	1292	-1.81	3.95
Pākehā	0.91	1.10	0.82	0.41	964	-1.35	3.00
Background food programmes*							
Targeted (v none)	-2.34	3.13	-0.75	0.46	60	-9.12	2.70
Universal (v none)	-4.04	3.56	-1.13	0.26	68	-11.75	2.56
School Model (Internal v External)	-1.80	1.63	-1.10	0.27	74	-4.84	1.47
Equity index	-0.59	1.25	-0.47	0.64	75	-2.87	1.90

Table 38: Net effect of the pilot's influence on measured change in mental wellbeing (WHO-5) while controlling for confounder variables using mixed modelling (n=1305).

						95% C.I.for B	
Influencing Factors (full model)	Beta	SE	Wald	p	df	lower	upper
(Intercept)	-3.80	3.92	-0.97	0.33	121	-11.78	4.16
treatment_control1	-0.87	1.39	-0.62	0.54	63	-3.73	1.87
demo_age_1_1	0.36	0.36	1.02	0.31	118	-0.39	1.08
demo_x2_maori1	0.10	1.14	0.09	0.93	987	-2.30	2.32

Random effect $ICC_{adj.}$(class) = 0.014. $R^2_{marg.}$ = 0.01, $R^2_{cond.}$ = 0.02.

Standardising net effects

To compare the net effects of the pilot across the different measures (availability/consumption, satiety, and wellbeing), standardised beta values were obtained. These values were then transformed onto a 5-star scale (including half values) for visual comparison, with the largest effect (vegetable consumption) representing 5 stars and the effects of the other parameters scored relative to this.

Table 39: Standardisied effects – A3

Influencing factors	Net effects (Beta)		
	Raw values	Standardisied	Stars /5
Vegetable consumption	0.92	0.839	5
Snack consumption (reduction in)	0.62	0.427	2.5
Fruit consumption	-0.02	0.073	0.5
Fullness	0.13	0.167	1
Total Health Quality of Life	2.97	0.220	1.5
Emotional construct	3.45	0.167	1
Physical construct	3.69	0.198	1
School construct	2.31	0.060	0.5
Social construct	2.31	0.117	0.5
Mental Wellbeing	-0.87	-0.044	-0.5

Appendix D: Learner food diary

Learner food diary

Rātaka kai mō te ākonga

Name: _____
Ingoa
Print name here
Āta tuhia tō ingoa ki konei

Class or Teacher: _____
Kaiako

Free and Healthy School Lunch Programme Evaluation

We need information to evaluate the Free and Healthy School Lunch Programme to make sure we give you the support you need. We would like to know about the what you ate, how full you feel, and how it makes you feel being at school over one week in September, and again over one week in November.

What if I don't want to complete the diary?

Participation is voluntary. If you don't want to answer the questions this won't affect you in any way.

Who is doing the evaluation?

Standard of Proof, an independent research and evaluation company, is doing the evaluation. They will keep your information confidential.

Are there any risks from my involvement?

There are no known risks. Your name won't be collected, so your response will be anonymous. After the evaluation ends in 2021, we will publish the findings. These won't identify any individual who took part.

Any questions?

If you have any questions about the evaluation contact Standard of Proof (hello@standardofproof.nz). You can also ask for a copy of the evaluation report after 2021.

Privacy Statement

- We're collecting information for the purposes of undertaking a formal evaluation of the Free and Healthy Food in Schools programme. We'll only use this information for evaluating the programme and to identify improvements and make decisions about the programme.
- The collection, storage and use of personal information will be in accordance with the Privacy Act 1993.
- Under the Privacy Act, individuals have the right to access any personal information that we hold about them, and can ask for it to be corrected. If you would like to make such a request, please contact hello@standardofproof.nz.

My first day
Taku rā tuatahi

1. Are you
He aha koe, he:

Boy *Tama*	Girl *Kōtiro*

2. Are you Māori?
He Māori koe?

Yes *Āe*	No *Kāore*

3. Are you a Pacific Islander?
He Pasifika koe?

Yes *Āe*	No *Kāore*

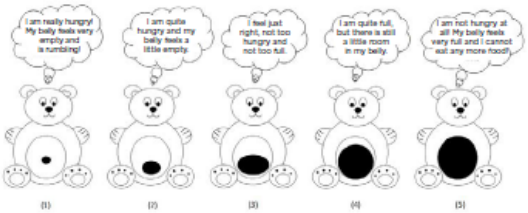
4. Are you New Zealand European?
He Pākehā koe nō Aotearoa?

Yes *Āe*	No *Kāore*

5. How old are you?
E hia ō tau?

5	6	7	8	9	10	11	12	13

6. How do you feel now? (Circle one teddy bear)[1]
E pēhea ana koe i tēnei wā? (Mākahia tētahi teti pea ki te porohita)



● - Food in Teddy's belly

If you are aged 8 or older, complete the next two questions (7 and 8). If you are aged 5, 6 or 7, you can stop now.
Mehemea e 8 ō tau neke atu rānei, tēnā whakakīa ngā whakautu e rua e whai ake nei (7 me te 8). Mehemea e 5, 6, e 7 rānei ō tau, kei te pai noa kia mutu koe i konei.

[1]Bennett,C. & Blissett,J. (2014). Measuring hunger and satiety in primary school children. Validation of a new picture rating scale. Appetite,78,40–48.

My first day
Taku rā tuatahi

7. Please indicate for each of the five statements which is closest to how you have been feeling over the last two weeks.

Tūtohutia mō ia o ngā tauākī e rima, ko tēhea e tata ana ki te whakaatu e pēhea ana ō whakaaro i ngā wiki e rua kua pahure ake nei.

	Over the last two weeks *I ngā wiki e rua kua pahure ake nei*	All of the time *I ngā wā katoa*	Most of the time *Te nuinga o te wā*	More than half of the time *Neke atu i te haurua o te wā*	Less than half of the time *Iti iho i te haurua o te wā*	Some of the time *I ētahi wā*	At no time *Kāore kau*
1	I have felt cheerful and in good spirits *E koa ana au me te hihiko o te ngākau*	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0
2	I have felt calm and relaxed *He marie, he mākohakoha hoki aku whakaaro*	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0
3	I have felt active and vigorous *Kua ngangahau, kua pākahukahu hoki au*	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0
4	I woke up feeling fresh and rested *I taku ohonga kua hou ngā whakaaro, kua tā te manawa*	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0
5	My daily life has been filled with things that interest me *Kua whakakīia ia rā ki ngā mea ngahau*	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0

8. In the past ONE month, how much of a problem has this been for you...

	About My Health and Activities (problems with...)	Never	Almost Never	Some-times	Often	Almost Always
1	It is difficult for me to walk a few houses from home (about 100 metres)	0	1	2	3	4
2	It is difficult for me to run	0	1	2	3	4
3	It is difficult for me to play sport or do exercise	0	1	2	3	4
4	It is difficult for me to lift something heavy	0	1	2	3	4
5	It is difficult for me to have a bath or shower by myself	0	1	2	3	4
6	It is difficult for me to help around the house	0	1	2	3	4
7	I get aches and pains	0	1	2	3	4
8	I have low energy	0	1	2	3	4

	About My Feelings (problems with...)	Never	Almost Never	Some-times	Often	Almost Always
1	I feel afraid or scared	0	1	2	3	4
2	I feel sad	0	1	2	3	4
3	I feel angry	0	1	2	3	4
4	I have trouble sleeping	0	1	2	3	4
5	I worry about what will happen to me	0	1	2	3	4

	How I Get Along With Others (problems with...)	Never	Almost Never	Some-times	Often	Almost Always
1	I have trouble getting along with other kids	0	1	2	3	4
2	Other kids do not want to be my friend	0	1	2	3	4
3	Other kids tease me	0	1	2	3	4
4	I cannot do things that other kids my age can do	0	1	2	3	4
5	It is hard to keep up when I play with other kids	0	1	2	3	4

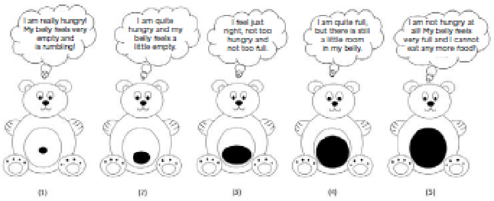
	About School (problems with...)	Never	Almost Never	Some-times	Often	Almost Always
1	It is hard to pay attention in class	0	1	2	3	4
2	I forget things	0	1	2	3	4
3	I have trouble keeping up with my school work	0	1	2	3	4
4	I am away from school because I feel sick	0	1	2	3	4
5	I am away from school to go to the doctor or hospital	0	1	2	3	4

Day 2
Rā 2

1. Today is
Ko tēnei rā ko te:

Monday *Rāhina*	Tuesday *Rātū*	Wednesday *Rāapa*	Thursday *Rāpare*	Friday *Rāmere*

2. How do you feel now? (Circle one teddy bear)[1]
E pēhea ana koe i tēnei wā? (Mākahia tētahi teti pea ki te porohita)



- Food in Teddy's belly

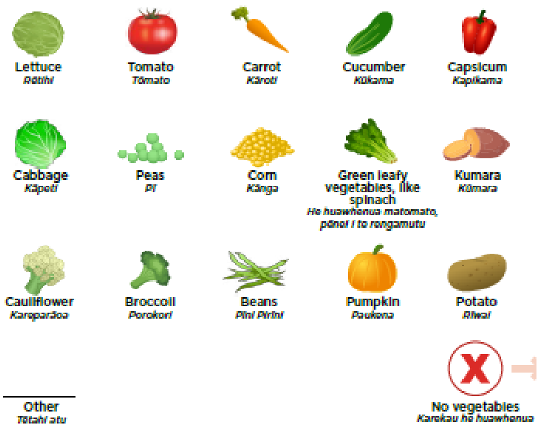
3. Did you eat lunch today?
I kai koe i te tina i tēnei rā?

Yes *Āe*	No *Kāore*

[1] Bennett, C. & Blissett, J. (2014). Measuring hunger and satiety in primary school children. Validation of a new picture rating scale. Appetite, 78, 40–48.

VEGETABLES
Ngā huawhenua

5. Circle any of the following vegetables that were in your lunch today. This includes everything that was in your lunch, no matter who gave it to you.
Tēnā porowhitatia ērā o ngā huawhenua e whai ake nei i kitea i roto i tō tina i tēnei rā. Ka uru ki tēnei huinga ngā mea katoa i roto i tō tina, ahakoa nā wai i homai ki a koe.



6. Make a big X on any of the vegetables you circled BUT DID <u>NOT EAT</u>
O ngā huawhenua i porohitatia e koe i runga ake nei, whakakorea ērā (ki tētahi X nui) huawhenua kāore i kainga e koe i te rā nei

Day 2
Rā 2

SNACKS AND SWEETS
Ngā paramanawa me ngā rare

7. Circle any of the following snack and sweet foods that were in your lunch today. This includes everything that was in your lunch, no matter who gave it to you.
Tēnā porowhitatia ērā o ngā paramanawa me ngā kai reka e whai ake nei i kitea i roto i tō tina i tēnei rā. Ka uru ki tēnei huinga ngā mea katoa i roto i tō tina, ahakoa nā wai i homai ki a koe.


Potato crisps, chippies or similar (burger rings, rashuns, twisties)
He rīwai pakapaka, he tiepi, me ērā momo (ngā burger rings, ngā rashuns me ngā twisties)


Fruit or muesli bars, or bliss balls
He ngaungau muesli, huarākau rānei, he bliss ball rānei


Corn chips
Ngā tiepi kānga


Popcorn
Kānga pāhihī


Lollies or gummies
He rare, he gummies, he candy rānei


Crackers
Ngā crackers


Cookies or sweet biscuits
He pihikete reka


Chocolate
He ngaungau tiakarete


Chocolate- or sugar-coated snacks
he paramanawa, he tiakarete, he huka rānei te uhi


Cake or muffin
He keke he māwhene rānei

Other
Tētahi atu


No snacks or sweets foods
Karekau he paramanawa, he kai reka rānei

8. Make a big X on any of the snacks and sweets you circled BUT DID NOT EAT
O ngā rare me ngā paramanawa i porohitatia e koe i runga ake nei, whakakorea ērā (ki tētahi X nui) kāore i kainga e koe i te rā nei

FRESH FRUIT
Ngā huarākau

9. Circle any of the following fresh fruit that were in your lunch today. This includes everything that was in your lunch, no matter who gave it to you.
Tēnā porowhitatia ērā o ngā huarākau e whai ake nei i kitea i roto i tō tina i tēnei rā. Ka uru ki tēnei huinga ngā mea katoa i roto i tō tina, ahakoa nā wai i homai ki a koe.


Banana
Panana


Apple
Āporo


Orange, mandarin or tangelo
He ārani, he mandarin, he tangelo rānei


Pear
He Pea


Grapes
Ngā Karepe


Pineapple


Berries
He Patatini


Kiwifruit
He Huakiwi


Stonefruit, like peaches or plums
He huarākau whai whatu, pēnei i te pītiti, i te paramu rānei


Watermelon or melon
Wātamerengi, he merengi rānei

Other
Tētahi atu


No fresh fruit
Karekau he huarākau

10. Make a big X on any of the fresh fruit you circled BUT DID NOT EAT
O ngā huarākau i porohitatia e koe i runga ake nei, whakakorea ērā (ki tētahi X nui) kāore i kainga e koe i te rā nei

Thank you very much
Ngā mihi nui!

Appendix D: School lunch self-assessments tool

Instructions

We are asking providers to complete a lunch assessment. We ask that each provider completes the five (green) tabs on this spreadsheet, making clear the ingredients in a standard lunch option provided to learners from 16-20 November. For each day, that day, we ask that you list the ingredients for a lunch portion provided to children.

What if we provide multiple lunch options for this school?

Please chose **one *standard* lunch** at the school, but ensure it includes all of the elements given to a child as a portion.

What if we provide lunches to multiple schools?

Please provide a lunch assessment for each school.

What do we do with this form once completed.

Please send this password protected file to: **team@standardofproof.nz**

Example day

Step 1. List standard menu for one child for the day.

Step 2. List ingredients that make up the a school lunch, e.g. list a sandwich as "white bread", "ham", "cheese slice", along with other elements of the lunch.

Step 3. For each ingredient, select the appropriate category from the drop-down list, e.g. "bread and crackers", "nuts and seeds"

Step 4. Select a rating - 'Green', 'Amber' or 'Red' - for each ingredient, using the **Rating Guide**.

Step 5. Estimate the percentage of the school lunch made up by each ingredient, ideally using the weight of each item if you know it. Note: the total ***should not exceed 100%***

Lunch	Ingredient list	Category	Rating	% of Lunch
				100%
Chicken sandwich, with fruit and muffin	Wholemeal bread	Breads and crackers	Green	20%
	salad greens	Vegetables and fruit	Green	25%
	lite mayo	Fats, oils, and spreads	Green	5%
	shredded chicken (no skin)	Poultry (chicken and turkey)	Green	30%
	apple	Vegetables and fruit	Green	10%
	Wholemeal bluberry muffin (prepackaged)	Baking	Amber	10%

Appendix E: Focus group interview schedule

1. Can you confirm when your school started/will start the programme?
2. What changes have you seen in your school over the past three months (since September)?
    a. What do you believe specifically contributed to this change? How do you know?
    b. Could there have been other factors contributing to this change?
        i. Probe to: food security; attendance; mental and physical health; academic achievement.
3. What changes have you seen in your community over the past three months (since September)?
    a. What do you believe specifically contributed to this change? How do you know?
    b. Could there have been other factors contributing to this change?
        i. Probe to: connections between the school and community; family health; food security in households; businesses in the region.

Appendix F: Evidence quality standards

SUFFICIENCY

An evaluation can focus on identifying new information, exploring context and meaning within the evaluand. Other evaluations include examining context through demonstrating differences between groups or identifying factors relevant to an evaluand. Whatever the purpose, it is important that the reader understand how adequately the data reflect the context relevant to the evaluation, and whether you have sufficient data to detect difference or understand the extent of relevant factors.

What can we say about the sufficiency of the data relevant to the context?	Data are sufficient in size to address the specified purpose	Data are sufficient to address the specified purpose, with limitations	Data are informative. There will be notable gaps in data, and more or other data will be required to understand the context
How do we know?	*Sample design relevant to the context, and data has achieved saturation, or demonstrates sufficient power to describe between-group differences or for testing relationships between relevant variables.* *e.g. saturation and/or sufficient and appropriate statistical power in quantitative approaches*	*Relevant sample design and data, but there are likely gaps in data relevant to learn about the context.* *e.g. saturation beginning, and/or insufficient statistical power to detect likely or possible effect sizes.*	*Data are too few to explore contexts relevant to the given purpose.* *e.g. saturation unlikely to have been achieved; statistical power not adequate at any level.*

GENERALISABILITY

Understanding the generalisability provides readers with clarity on how to use (and not use) the evidence as a reflection of evaluand.

What can we say about the generalisability of the results?	Data are generalisable to the relevant population	Data have some degree of generalisability, with limitations	Data are informative but not likely generalisable; more other data are required to generalise the results
How do we know?	*Balanced sample design and data, and sampling approach and response rates allow results to be generalised to the population of interest.* *e.g. census, probability sampling approaches using complete sample frames and acceptable margin of errors, accounting for the design effects.*	*Balanced sample design and data. Results can be generalised to groups relevant to the evaluand, and/or there is likely to be gaps in the ability to generalise (i.e. sample frame limitations, large margin of errors, design effects not adjusted).*	*Balanced sample design and data, fit for the given purpose. Data have not been sampled to allow for generalisability, considering design, sample frames, resulting data and response rates.* *e.g. purposive sampling only.*

BALANCED

An unbalanced approach or tools can lead to overestimating or underestimating the effects or influences relevant to the evaluand, or reflect specific, prioritised groups or experiences. Balanced and/or unbalanced designs may be the goal, which is why such information is vital for readers to understand.

Are the data presenting a balanced view of the evaluand?	Results are likely a balanced depiction of the evaluand, with low risk of inaccuracy	Results may be a balanced depiction of the evaluand, although moderate risk of bias exists	Data are informative and prioritised to specific experiences; other data are required to provide a balanced picture of the evaluand

How do we know?	*Sample design and approach mitigates and confirms low risk to potential bias in what is being measured (e.g. selection and attrition, measurement, performance bias) and unbalanced results. The data likely highlight the phenomena of interest from the participant's eyes considering the full evaluand.* *e.g. standardised assessments, formal assessments of validity and/or reliability, psychometric analysis of items and instruments, general elimination methods, validation techniques.*	*Sample design and data collection approach focuses on eliciting unique perspectives. Techniques to confirm validity of what is being measured not possible or completed OR techniques have shown likely bias.*	*Sample design is biased towards specific experiences (e.g. advocacy or strengths-based approach, selective reporting)*

ADEQUATE COLLECTIVE EVIDENCE

Using adequate data, and a mixture of methods and data sources provides ample opportunity to test theories, views, experiences and achievements, each providing potentially unique realities relevant to the evaluand. Triangulating evidence promotes a balanced picture, supporting the **credibility of evaluation** through the overall weight of the evidence.

Are the evidence sufficient to answer the evaluation questions?	Evidence are sufficient to the answer the question	Evidence are somewhat sufficient to answer the evaluation question	Evidence are informative and may reflect select experiences
How do we know?	*The key methods used to answer the question are sufficient in size and/or power to fully address the question, while multiple data are triangulated to provide greater insight and test the findings and any potential bias in data.*	*Although no method is sufficient in size relevant to the context and question, at least one method is somewhat sufficient. Triangulation used across multiple data sources, although there exists potential bias in the methods,*	*Few methods and data sources used, and/or no triangulation giving a limited picture of weight and/or balance of evidence.* *e.g. single methods/dataset*


STANDARD OF PROOF
Evaluation Research Strategy