CHAPTER 4:MAIN STUDY METHODS

Chapter overview

4.1 SAMPLING

4.1.1 Recruitment, characteristics and time frame

All learners of three of the five grade six classes of a middle-class, predominantly white, Afrikaans medium public primary school in Pretoria, South Africa, were chosen. These three classes were taught by the same mathematics teacher (and the food recording part of the validation was in the context of mathematics). This was the same school (but different children) that had been used for the developmental evaluation of reference method 1 in the previous year. (The developmental evaluation sub-studies on the test method as described in the previous chapter were done in another, but demographically comparable school.) The aim was to achieve at least a sensitivity of 80% and a specificity of 60% for the screener. With 27 positive and 27 negative responses, significant results would be obtained at a 95% confidence level and a power of 80%.

For the test-retest reproducibility assessment, a random sub-group of 13 children was selected from each of the three classes.

An overview of the time frame and all the stages, including the developmental evaluation discussed in the previous chapter and details regarding the data collection of each stage are provided in Table 4.2.

4.1.2 Ethical approval and consent

The Research and Ethics Committee of the University of Pretoria approved the project (Protocol 4/2000). Following an information meeting, permission to do the study was obtained from the headmaster of the school, who also informed the relevant authorities. The deputy headmaster, who was also the mathematics teacher, handled practical matters. A joint letter from the researcher and the headmaster explaining the study, and in particular how and why it was integrated into the mathematics curriculum, was sent to all learners in the identified classes. Informed and willing parental consent and agreement of the children in line with published guidelines was obtained.[269] (Addendum E)

No incentives, apart from a (unannounced) pen and a snack following the data collection, were offered.

4.2 DATA COLLECTION

4.2.1 Test method

Each of the three classes was divided into three groups of about 12 children each. Data collection was done in this group context and was fitted into the school timetable. The first administration took place in the beginning of September 2001. The second administration (test-retest reproducibility study) followed on average six weeks later. Total contact time per administration per group was about 45 minutes.

4.2.1.1 Setting

The conference room of the school was used for data collection. Up to twelve learners were seated in a continuous U-shape with separators placed on the table between adjacent participants to ensure privacy of response. On arrival of the children each 'booth' contained an answer sheet (Addendum F), a coloured cover sheet (positioned to guide the learners to code their response correctly) and a pen. A poster size version of the answer sheet was stuck on the front wall. This was also fitted with a cover sheet similar to the one of the respondents. An overhead projector, the transparencies, the flip-file, the portion size estimation aids and pointers completed the setting.

4.2.1.2 The interview

All data collection regarding the test method was done by the researcher personally. After the introduction participants were reminded that the project involved research, that their responses were a confidential and private matter (hence the separators), and that truthful answers reflecting their typical ('normal') eating habits since the beginning of the year should be reported (that is "since you were in grade six").

One ground rule was set, namely that no value-laden comments about food would be allowed. Clarifying questions, were, however, encouraged. Throughout a session great care was taken to maintain a friendly, relaxed atmosphere, yet restricting discussion to clarifying questions. The coded nature of the answer sheet ensured that participants within a group proceeded at the same pace.

Before commencing, the interviewer made sure that the cover sheet was positioned properly on the answer sheet (that is below the first row [M], and instructed participants to keep the sheet exactly as the example on the poster. This ensured that responses were coded at the appropriate spaces of the answer sheet and missing data would be minimised.

For each food category, an explanatory sentence was made when the full-colour composite photo of the relevant foods was projected on a screen and a separate, identical flip file version was turned to the same page (for example "This is a picture of ……"). The exact text was written on the reverse side of the flip file, facing the interviewer, and was provided in the mother tongue of the children, Afrikaans. Care was taken to highlight distinguishing feature(s) of the category by pointing to it on the picture (for example *full fat / whole* milk, "*paper-wrapped*" margarines, "*lite / diet / fat free*" labels etc) and to always mention all foods included in the category. The interview started with: "Do you eat foods such as those on the picture?" Correct answering technique was demonstrated on the poster replica of the answer sheet. Participants who responded "no" were requested to put down their pens and only proceed when instructed to do so. (This filter question took extra time, but during piloting proved to provide clarity and eliminate confusion and ambiguous responses later on.)

Typical frequency of intake was requested next: "If you eat food like those on the picture every day, write down the number of times you usually eat it during one day on the line saying 'per day'. If you do not eat such food every day, move to the 'per week' line, and fill in there how often in the course of a week you usually eat such foods. Only write in the 'per day' or the 'per week' line." In the case of eggs only a 'per week' option was given, and in the case of table fats only a 'per day' option was provided. The option was given to write <1 in the 'per week line'. Equally, if children felt that their usual consumption was within a range (for example 2 to 3 times per week) this could be indicated (and was then coded, for example 2.5). A practical example of how to answer this was given: "Jannie usually has bacon or a vienna sausage as part of his breakfast, a ham sandwich for lunch and, for example, a chop for supper. He thus usually has meat three times per day. If you were Jannie, you should write a 3 in the per day line".

The last question for each food category was about usual amount consumed. For this purpose a combination of 2D PSEA (geometric shapes, for example a 90mm diameter circle for meat), household measures (cups and spoons, for example a 250ml measuring cup for milk) and photographs (for example chocolate, nuts, chips and high fat crackers) (see Table 3.2) were used to give a visual indication of the reference serving. Children were instructed to mark "2" if the amount usually consumed was similar, "1" if it was about half as much, and "3" if it was one and a half times as much as the reference amount.

The above procedure was repeated for all the food categories.

4.2.1.3 Test-retest reproducibility

The second administration (test-retest reproducibility study) was conducted in exactly the same standardised way as described above on average six weeks later.

4.2.2 Reference method 1: Food record

Training and data collection for the three-day food record were done according to the protocol established during the testing phase (see previous chapter). Each of the three classes was in a specific recording group. An extra (mixed) group was formed of children from these three classes who did not participate in a class tour (September 2001; midweek recording group, that is group 2), since the teachers considered this a meaningful activity. In this way a greater percentage of children used electronic scales, even though the midweek recording group (who did not have a weekend day) became proportionally larger.

In Table 4.1 the programme is summarised, showing that for the group as a whole all days of the week were represented and that the last day of recording fell on different days of the week. In one group a weekend day was the first day of recording. Per group 16 randomly chosen children performed the weighing using a supplied electronic scale (in total 64 of the children). The rest either used their own (spring) scales or were supplied with a set of measuring utensils (spoons, cups and ruler). The accuracy of the own scales was not checked.

TABLE 4.1: FOOD RECORDING AND TRAINING PROGRAMME FOR REFERENCE METHOD 1

Recording group	Training	Recording days	Hand-in
1	Wed 10/10/2001	Thu 11/10/2001 Fri 12/10/2001 Sat 13/10/2001	Mon 15/10/2001
2	Mon 15/10/2001	Tue 16/10/2001 Wed 17/10/2001 Thur 18/10/2001	Fri 19/10/2001
3	Fri 19/10/2001	Sun 21/10/2001 Mon 22/10/2001 Tue 23/10/2001	Wed 24/10/2001
2 (mixed class)	Mon 17/9/2001	Tue 18/9/2001 Wed 19/9/2001 Thur 20/9/2001	Fri 21/9/2001

4.2.3 Reference method 2: Screener by parents

The dietary fat screener, which was to be completed by the parents in respect of their grade six child, was sent to them together with the information letter and the informed consent. A direct caregiver was requested to complete the screener and return it with the child to the mathematics teacher from whom the researcher collected it. One week after the initial handing-out, children were requested to write a reminder in their homework books.

TABLE 4.2: OVERVIEW OF STAGES AND LOGISTICS OF DATA COLLECTION

What?		Why? (Aim)	How?	When (where)?	Data-set
Development and developmental evaluation	Initial development of test method	Target group specific dietary fat screener (specific criteria in text)	Content experts designed a novel South African version of the NCEP's MEDFICTS • Item list: Pictures; South African foods eaten by target group • Portion size estimation aid: Two-dimensional graphics and photos of typical foods • Graphic depiction of frequency of intake	1999	-
	Developmental evaluation of test method	Content / face / consensual validity Construct / measurement validity	Testing among separate groups of respondents some of the concepts contained in dietary fat screener, i.e. • Item list (sub-study 1) • Reference portion size (sub-study 2) • Portion size estimation aids (sub-study 3) • Depiction of frequency of intake (sub-study 4)	2001 2000 (School MP for sub-studies 2-4)	D1 D2 D3 D4
	Developmental evaluation of reference method 1	Piloting and standardisation of reference method 1 Content validity of test method	Testing of three-day weighed food record (sub-study 5) All grade 6 learners complete three consecutive days food record using supplied electronic scales, own scales or household measures; part of mathematics assignment (see R1-2)	2000 (School SR)	R1-1
Main study	Informed consent	Ethical conduct Permission	Headmaster, educator, parents and learners Information letters and consent forms handed out to learners in school	2001 Week 1 (School SR)	-
	Administration of reference method 2	Concurrent / criterion validity	Parental completion of screener Screener in self-completion, text format included in information/consent package sent with learners to parents	2001 Week 1 (School SR)	R2
	Collection of anthropometric and biographic data	Concurrent / predictive validity Description of sample	Weight and height of children measured by teacher as part of mathematics class; apparatus [Tanita scale and mobile height gauge] supplied by researcher; teacher trained for standardised technique regarding subject clothing and recording [to nearest 0.1kg or 0.01m for weight and height respectively]; attention to privacy/confidentiality during data collection and recording	2001 Week 1 (School SR)	A
	First administration of test method	Obtaining test data Assessment of statistical properties of test method	Screener completed by children: Nine groups each ± twelve learners; All researcher-administered in school-time using set procedure which includes physical setting [school conference room with a furniture arrangement simultaneously conducive to interviewer-subject interaction, subject privacy and interviewer control in terms of completeness and appropriateness of response on the answering sheets etc], use of visual aids, pacing, anticipated guidance etc	2001 Week 2 (School SR)	T1
	Administration of reference method 1	Criterion / construct (convergent) validity Content validity	Three-day weighed food record by children: Four groups from three classes (each ±36 learners) weigh and record all food intakes each for 3 consecutive days as a mathematics assignment. Weighing equipment [Electronic Soehnle scales and household measuring cups and spoons and a ruler], colour-coded recording forms, flow-diagrams to assist with food description] supplied by researcher, as well as a reference file and a demonstration kit for the training session [regarding proper use of measuring equipments and recording technique]. Group 1: (Wednesday: training), Thursday, Friday, Saturday recording (Monday handing in) Group 2: (Monday: training), Tuesday, Wednesday, Thursday recording (Friday handing in) Group 3: (Friday: training), Sunday, Monday, Tuesday recording (Wednesday handing in)	2001 Week 4-6 (School SR)	R1-2
	Second administration of test method	Test-retest reproducibility	Screener completed again by children: Three groups of each 13 randomly selected learners from the three classes; all researcher-administered in school-time; procedure as in first administration (T1)	2001 Week 8 (School SR)	T2

4.2.4 Anthropometric data

Weight and height were obtained as part of mathematics activities using standard techniques,[26] except that children were dressed in summer school uniform (no footwear and jerseys). Weight measurements were taken accurate to 100g and height to the nearest 0.1cm. All measurements were taken in the mornings (between 08:00 and 10:00, mid September 2001) by the same teacher in a private corner of the classroom. Equipment (Tanita electronic scale [Tokyo] and portable height gauge), formal training on proper technique and recording, as well as data collection forms were provided to the teacher. Date of birth was obtained from school records. Privacy and confidentiality were high priority.

4.3 DATA PROCESSING AND ANALYSIS

Data cleaning and coding into EXCEL of the test method, reference method 2 and anthropometric data were done by the researcher personally. The EXCEL spreadsheets were imported to SAS (mainframe version 8.2), where all analyses were performed, except for Kappa, McNemar (for three by three tables) and Friedman statistics, which were done on BMDP statistical software release 7.1. Input of all data was checked by the researcher personally. Programming was done by a professional programmer in consultation with the Statistical Advice Center (STATOMET) of the University of Pretoria.

4.3.1 Description of sample

For the anthropometric description of the participants, the CDC 2000 growth data files for boys and girls aged two to 20 years and the accompanying SAS software were used for describing mean age, weight, height, body mass index (BMI, in kg/m^2) as well as weight for age, height for age and BMI for age in terms of mean centiles and Z-scores (http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/datafiles.htm, accessed 9/12/2001). Current age was calculated in months based on the actual date on first assessment and date of birth.

4.3.2 Test method

4.3.2.1 Scoring

The steps in the scoring process were as follows:

- If frequency of intake was reported as daily consumption, this was converted to weekly consumption by multiplication by seven.
- Weekly consumption was categorised and scored as specified in the original tool:

- Less than once per week was scored zero
- Once or more (up to three times per week) scored three points
- More than three times per week was scored seven.

- If non-consumption was reported for a food category, the weekly consumption and the portion size were assigned the value zero.
- Category scores were calculated as in the original MEDFICTS tool, that is by multiplying the weekly consumption score with the portion size score (that is 1, 2 or 3).[203]
- All ten category scores were added to create a final score using the SAS assignment statement (in contrast to the sum function) in order to ensure that missing values in either the frequency of intake score or the portion size score would result in a missing final score. (This was done to prevent final scores from reflecting less than ten category scores and thus indicating an erroneous low final score.)
- The final score, which could range from zero to 210, was categorised as 'high fat' if it was more than 68. A final score of less than or equal to 68 was classified as 'prudent'.

4.3.2.2 Internal consistency

In order to explore the test method following classical test theory, the following analyses were performed:

- Item total correlations (Pearson) between all ten category scores and the final scores
- Cronbach's coefficient alpha
- Split half method, whereby the ten food categories were randomly assigned to two groups and Pearson's correlation coefficient was calculated between the groups.

4.3.2.3 Test-retest reproducibility

The *test-retest reproducibility* was determined as follows:

- A check for sampling bias was performed using Wilcoxon's Rank Sum Test to assess (within the first administration) whether children included in the re-test differed significantly in respect of their category and final scores from those who were not re-tested.
- The degree of agreement for the portion size and frequency of intake estimations in the two administrations was expressed as percentage of pairs with exact (=identical) agreement.
- The kappa statistic was used to estimate chance-corrected proportional agreement. It was interpreted according to the guidelines suggested by Altman.[270]

- McNemar's statistic of symmetry was used to test for equality of frequencies in all pairs that were symmetric around the diagonal of perfect agreement.
- The linear relationship between the final scores in the test and the re-test was measured by means of the Spearman correlation coefficient ('reproducibility correlation').
- Indicators of random error (variability) such as standard deviations and confidence intervals (95%) were calculated.
- Wilcoxon Signed Rank Test was used to assess the significance of the difference between the first and the second administration regarding the ten category scores and the final scores.
- The differences between the final scores in the two administrations were plotted against the mean of the two scores (Bland Altman method).[270]

4.3.3 Reference method 1: Food record

The three-day food records were analyzed by an experienced registered dietitian using FoodFinder3®, the most current food database of the Institute for Nutrition Intervention Research of the Medical Research Council (MRC) of South Africa. The dietitian was of the same culture as the target group and familiar with Afrikaans children's eating habits, language usage and trends in the food industry.

Before the coding commenced the researcher and the coder together laid down a number of 'coding rules'. These were updated as needed. Ongoing consultation was maintained during the coding phase and if assumptions had to be made, these were a joint decision. The researcher herself checked every record for the following: Choice of food item from the database, comprehensiveness of coding (all items entered) and correctness of amounts. Editing was done by the researcher, whereafter data were exported to EXCEL and imported into SAS.

The following steps were followed to obtain the measures of high fat intake:

- Mean daily energy (kJ), total fat (g), saturated fatty acid (g) and cholesterol (mg) intakes over the three days were calculated for each participant.
- Mean total fat and saturated fatty acid intakes were converted to energy (kJ) equivalents by multiplication by 37.8.
- PFE and PSFE were then calculated by expressing total fat energy and total saturated fatty acid energy as a percentage of mean daily energy intakes.
- The food record information was classified as 'high fat' when:
    - PFE > 30

- PSFE > 10
- Mean daily cholesterol intake => 300mg.

- Conversely the diets were classified as 'prudent'.
- A variable 'ANY' was created to indicate that any one of the three measures reflected high fat intake.
- The variable 'ALL' meant that all three conditions were met simultaneously.

(Thus eventually five outcome measures of high fat intake were created.)

4.3.3.1 Data quality assurance

Mean reported energy intake was evaluated against presumed energy requirements. In order to estimate the latter, the basic metabolic rate (BMR) was calculated by using the WHO formula (FAO/WHO/UNU, 1985): [271]

Girls (10-18 years): BMR(kJ) = [12.2* Weight) + 746]* 4.2

Boys (10-18 years): BMR(kJ) = [17.5*Weight) + 651]* 4.2

The Physical Activity Level (PAL) was calculated as the ratio of mean daily energy intake to BMR.

Mean daily energy intake was expressed as a percentage of the 2002 Dietary Reference Intakes (DRI) for children nine to 13 years, that is 9572kJ and 8698kJ for boys and girls respectively (active PAL).[10]

Correlations (Pearson) were calculated between mean energy intake over the three days and weight and BMI for the whole group and for genders separately.

Mean recorded energy intake between the following sub-groups was compared:

- Electronic scale users versus estimation with household measures
- First versus second versus third day of recording (check for recording fatigue)
- Recording period one versus two versus three (weekday versus weekend day effect)

The latter two were combined in a cross frequency and the Friedman two-way analysis of variance with multiple comparisons was computed.

4.3.4 Reference method 2: Screener by parents

Scoring and analysis of internal consistency of the screener as completed by the parents (reference method 2) were done in the same manner as for the test method (see above).

4.3.5 Comparison of test method to reference methods

4.3.5.1 Test method versus food record

In order to compare the test method to the food record the following statistical analyses were performed:

- Spearman rank correlation coefficients between the final score obtained in the test method and the measures of fat intake from the three-day food record (that is total fat, PFE, saturated fatty acids, PSFE and cholesterol) were calculated.
- A multiple two by two table between, on the one hand, the classified final score of the screener, and, on the other hand, the five dichotomised outcome measures from the three-day food record (PFE, PSFE, cholesterol, 'ANY' and 'ALL') was set up to illustrate (percentage) classification agreement.
- Based on the mentioned tables chance corrected agreement (simple kappa) was calculated.
- The indicators of comparative validity presented in Table 4.3, were determined with 'high fat' denoting 'positive' and 'prudent' meaning 'negative'.

TABLE 4.3: INDICATORS OF COMPARATIVE VALIDITY AND THEIR FORMULAE

Indicator	Description	Formula[a]
Sensitivity	Proportion of individuals with high fat intake who were correctly identified by the screener as being at risk	TP / (TP+FN)
Specificity	Proportion of individuals following prudent diet correctly classified by the screener as not at risk	TN / (TN+FP)
Overall predictive value	Proportion of predictions that are true positives and negatives	(TP+TN) / Number of predictions
Positive predictive value	Proportion of positive tests that are true	TP / (TP+FP)
Negative predictive value	Proportion of negative tests that are true	TN / (TN+FN)
Relative risk	Ratio of two incidence rates	$\frac{TP / (TP + FP)}{FN / (TN+FN)}$
Odds ratio	Ratio of four incidence rates	$\frac{TP \times TN}{FP \times FN}$

[a]TP = True positives
TN = True negatives
FN = False negatives
FP = False positives

Mean intakes (energy, fat, PFE, saturated fatty acids, PSFE and cholesterol) consumed by those classified as high fat and prudent by the test method.

4.3.5.2 Test method versus screener by parents

The statistics performed to compare the test method to the screener as completed by the parents involved the following:

- The degree of agreement for the portion size and categorised frequency of intake estimations between children and parents was expressed as percentage of pairs with perfect (=identical) agreement.
- The kappa statistic was used to estimate chance-corrected proportional agreement of categorical variables (portion size, categorised frequency of intake, categorised final score). It was interpreted according to the guidelines suggested by Altman.[270]
- For categorical variables (portion size, categorised frequency of intake, categorised final score) McNemar's statistic of symmetry was used to test for equality of frequencies in all pairs that were symmetric around the diagonal of perfect agreement.
- Mean category and final scores were calculated.
- The linear relationship between the category and final scores of the children and parents was measured by means of the Spearman correlation coefficient for the whole group and for the genders separately.
- Indicators of random error (variability) such as standard deviations and confidence intervals (95%) were calculated.
- The difference between children's and parents' final scores was calculated for the whole group and for the genders separately.
- Wilcoxon's Signed Rank Test was used to assess the significance of the difference between children's and parents'final scores.
- The differences between the final scores of child and parent pairs were plotted against the mean of the two scores (Bland Altman method).[270]

4.3.5.3 Test method versus both reference methods

The amount of agreement among classifications as derived simultaneously by the test method, reference method one (all five outcome measures) and reference method two was determined and each triangulation was graphically depicted using Venn diagrams.

4.3.5.4 Receiver operating characteristics

Based on a linear regression model, receiver operating characteristics (ROC) curves were created by calculating for each of the five outcome measures (PFE, PSFE, cholesterol, 'ANY' and 'ALL') from reference method one, the sensitivity and specificity of every observed final score

and plotting the sensitivity against 1-specificity. The area under the ROC curve was determined as a global assessment of the discriminatory performance of the screener relative to each of the outcome measures.