DEVELOPMENT AND COMPARATIVE VALIDATION OF A DIETARY FAT SCREENER FOR GRADE SIX CHILDREN

by

FRIEDEBURG ANNA MARIA WENHOLD (neé Pape)

Submitted in fulfillment of the requirements of the degree

Philosophiae Doctor

in the

FACULTY OF HEALTH SCIENCES

UNIVERSITY OF PRETORIA

PRETORIA

SUPERVISOR: PROF DR P RHEEDER

CO-SUPERVISOR: PROF DR UE MacINTYRE

2004

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ADDENDA

ABSTRACT:

DEVELOPMENT AND COMPARATIVE VALIDATION OF A DIETARY FAT SCREENER FOR GRADE SIX CHILDREN

Friedeburg Anna Maria Wenhold
Supervisor: Prof Dr P Rheeder
Co-supervisor: Prof Dr UE MacIntyre
Division Human Nutrition, University of Pretoria
Degree: Philosophiae Doctor

Background

Risk factors for chronic non-communicable diseases have been shown to track from childhood into adulthood. Cost-effective intervention starts with valid screening. The aim of this research was development and comparative validation of a dietary fat screener in grade six learners.

Methods

A pictorial, quantitative food frequency questionnaire type, scored dietary fat screener (test method), consisting of ten food categories associated with high fat intakes, was developed and subjected to developmental evaluations in the target group. Subsequently the test method was administered to learners of an urban middle-class school (Pretoria, South Africa). Test-retest reproducibility was checked in a random sub-sample. Two reference methods were used for comparison: Parental completion of the screener and a three-day food record by the children.

Reliability testing of the test method involved measuring internal consistency and test-retest reproducibility. Credibility of energy intakes in the food record was checked. Mean cholesterol intake and percentage fat and saturated fat energy were determined. Comparative validation was based on correlations, mean differences and the Bland Altman method for continuous variables. Percentage agreement, kappa statistics and the McNemar tests were determined for categorical data, as were sensitivity, specificity and predictive values. Receiver operating characteristic curves were plotted.

Results

Sample: Out of 108 children, 39 (100%) were re-tested, 93 (86%) provided usable food records and 78 (72%) parents responded. Mean age was 148±4.4 months.

Reliability: The test method was internally consistent. Test-retest reproducibility of portion size and frequency of intake estimates depended on the food category. No systematic error between administrations was noted as mean category and final score differences between the two administrations did not differ significantly from zero. A significant (r=0.36, P=0.02) correlation existed between administrations, but boys were characterised by random error and a lack of reproducibility (r=0.26, P=0.29), whilst for girls reproducibility could be established (r=0.58, P=0.01).

Comparison to screener by parents: Children and parents did not agree in respect of reported portion size and frequency of intake. Parents had lower values for all scores. Correlation between children's and parents' final scores was 0.23 (P=0.04) (boys: r=0.13, P=0.46; girls: r=0.33, P=0.04), but the mean difference in final scores differed significantly from zero (P=0.0001). Classification was identical in 74% of cases, but when corrected for chance this agreement was also poor.

Comparison to food record: The food record appeared to be a plausible reflection of energy intakes during the recording period. For girls a significant (P<0.05) correlation between test method final score versus cholesterol intake and energy from total and saturated fat was found. The sensitivity of the test method was very high (over 90%). Chance corrected agreement between test method classification and measures of fat intake from the food record was poor. Changing the cut-off of the test method final score could not achieve high sensitivity and high specificity simultaneously.

Conclusion

The dietary fat screener cannot yet be used as sole indicator of high fat intake in grade six learners. Until the discriminatory abilities have been improved, its value lies in creating awareness of high fat intakes and providing a food-based starting point for anticipatory guidance.

Key words: Validity, reliability, reproducibility, nutritional assessment, screening, dietary fat, children, food record, sensitivity, agreement

OPSOMMING:

ONTWIKKELING EN VERGELYKENDE VALIDERING VAN 'N DIEETVET-SIFTINGSINSTRUMENT VIR GRAAD SES LEERDERS

Friedeburg Anna Maria Wenhold
Leier: Prof Dr P Rheeder
Medeleier: Prof Dr UE MacIntyre
Graad: PhD

Agtergrond

Risikofaktore vir latere ontwikkeling van chroniese nie-oordraagbare siektes word reeds in die kinderjare gevestig. Koste-effektiewe intervensie begin met geldige sifting. Die doel van hierdie studie was die ontwikkeling en vergelykende validering van 'n dieetvet-siftingsinstrument by graad ses leerders.

Metodes

'n Kleurprent-gebaseerde, kwantitatiewe voedselfrekwensievraelys-tipe dieetvet-siftingsinstrument (toetsmetode), bestaande uit tien voedselkategorieë met 'n bewese verband met hoë vetinnames, is ontwikkel en onderwerp aan 'n aantal ontwikkelingsstudies in die teikengroep. Daarna is die toetsmetode op leerders in 'n stedelike, middel-sosio-ekonomiese klas skool (Pretoria, Suid-Afrika) toegepas. Toets-hertoets-herhaalbaarheid is in 'n ewekansige sub-groep gekontroleer. Twee verwysingsmetodes is ter vergelyking gebruik: Voltooiing van die siftingsinstrument deur ouers en 'n drie-dag-voedselrekord deur die kinders.

Data-analise van die toetsmetode het betroubaarheidstoetsing in terme van interne, item-konsekwentheid en toets-hertoets-herhaalbaarheid behels. Geloofwaardigheid van gerapporteerde energie-inname in die voedselrekord is gekontroleer. Gemiddelde cholesterol-inname en persentasie energie van totale vet en versadigde vetsure is bepaal. Vergelykings is gebaseer op korrelasies, gemiddelde verskille en die Bland Altman metode vir kontinue veranderlikes. Persentasie ooreenstemming, kappa statistieke en McNemar toetse is bepaal vir kategoriese data, asook sensitiwiteit, spesifisiteit en voorspellingswaardes. "Receiver operating characteristic"-krommes is geplot.

Resultate

Steekproef: Van 108 leerders is 39 (100%) leerders hertoets, 93 (86%) het bruikbare voedselrekords verskaf, en 78 (72%) ouers het gerespondeer. Die gemiddelde ouderdom was 148±4.4 maande.

Betroubaarheid: Die toetsmetode was intern konsekwent. Toets-hertoets-herhaalbaarheid was item-afhanklik. Geen sistematiese fout is tussen die twee toedienings bevind nie, aangesien gemiddelde verskille in kategorie- en finaletellings tussen die twee toedienings nie betekenisvol van nul verskil het nie. 'n Betekenisvolle (r=0.36, P=0.02) korrelasie is gevind tussen finaletellings in die twee toedienings, maar seuns was gekenmerk deur ewekansige foute en 'n gebrek aan herhaalbaarheid (r=0.26, P=0.29), terwyl daar vir dogters herhaalbaarheid bevind is (r=0.58, P=0.01).

Vergelyking met sifting deur ouers: Kinders en ouers het nie ooreengestem ten opsigte van gerapporteerde porsiegrootte en frekwensie van inname nie. Ouers het laer waardes vir alle tellings gehad. Korrelasies tussen kinders en ouers se finaletellings was 0.23 (P=0.04) (seuns: r=0.13, P=0.46; dogters: r=0.33, P=0.04), maar die gemiddelde verskil in finaletellings het betekenisvol van nul verskil (P=0.0001). Klassifikasie was identies in 74% van gevalle, maar sodra dit vir toeval gekorrigeer is, was die ooreenstemming swak.

Vergelyking met voedselrekord: Die voedselrekord was geloofwaardig vir die rapporterings-periode. Vir dogters is 'n betekenisvolle (P<0.05) korrelasie tussen die toetsmetode se finaletellings en cholesterolinname asook totale energie vanaf dieetvet en versadigde vetsure bevind. Die sensitiwiteit van die toetsmetode was baie hoog (oor 90%). Toeval-gekorrigeerde ooreenstemming tussen die klassifikasie deur die toetsmetode en aanwysers van vetinname in die voedselrekord was swak. Verandering van die afsnywaarde van die toetsmetode se finaletelling kon nie hoë sensitiwiteit en hoë spesifisiteit gelyktydig bewerkstellig nie.

Gevolgtrekking

Die dieetvet-siftingsinstrument kan nog nie as alleenaanwyser van hoë vetinname gebruik word nie. Totdat die onderskeidingsvermoë verbeter is, lê die primêre waarde daarvan in bewus-making van hoë vetinnames en in die verskaffing van 'n voedsel-gebaseerde vertrekpunt vir voedingonderrig.

Sleutelterme: Geldigheid, betroubaarheid, herhaalbaarheid, voedingstatusevaluering, sifting, dieetvet, kinders, voedselrekord, sensitiwiteit, ooreenstemming

CHAPTER 1: RESEARCH PROBLEM IN CONTEXT

Chapter overview

1.1 RATIONALE FOR STUDY: THEORETICAL CONTEXT

The high prevalence of chronic non-communicable diseases (CNCD) such as coronary heart disease, diabetes mellitus and certain cancers, makes these conditions worldwide - in industrialised and developing countries [1] - as well as in South Africa [2] - public health issues. In fact, in the World Health Organisation (WHO) study on the Global Burden of Disease, ischaemic heart disease is on top of the list of causes of death worldwide [3] and is expected to stay there according to the 2020 projection.[4] King et al [5] estimate that by the year 2025 more than 75% of people with diabetes mellitus will reside in developing countries. This represents a 170% increase from current prevalence rates, compared to a projected increase of 42% in developed countries.

Whilst South Africa as a whole is classified as a developing country,[6] a major component of its population is in the so-called nutrition transition,[7] meaning that traditional lifestyles are increasingly replaced with Western eating habits.[8] The nutrition programming theory implies that developing societies may be subject to the double burden of disease, whereby early nutritional deprivation (such as maternal and fetal malnutrition commonly seen in developing countries) predisposes an individual to the development of CNCD's later in life.[9]

Numerous risk factors, including diet, are linked to CNCD as a group.[10] This has resulted in the formulation of dietary guidelines by various organisations, such as the National Heart, Lung, and Blood Institute's National (NHLBI) Cholesterol Education Program (NCEP), and the regularly revised and updated Dietary Guidelines 2000 of the American Heart Association[11] and the United States Department of Agriculture.[12] In South Africa a set of food-based dietary guidelines has been developed in accordance with international guidelines.[13] These country-specific, evidence based guidelines were officially approved and adopted by government in 2003 for use by South Africans seven years and older.[14] One of these guidelines stipulates: "Eat fats sparingly".

There is growing consensus that the dietary guidelines should include children in general terms, as well as specifically in respect of fat intake.[11, 12, 15, 16, 17] Not only did a meta-analysis of the NCEP Step 1 and 2 diets point to multiple beneficial effects on cardiovascular risk factors in adults,[18] but these dietary changes also resulted in lowering of low density lipoproteins (LDL) over three years while maintaining growth, iron stores, nutritional adequacy, and psychological well-being in children with elevated LDL-cholesterol concentrations.[19] Similarly, Obarzenek et al [20] reported from the Dietary Intervention Study in Children (DISC) trial on the feasibility, efficacy and safety of cholesterol lowering intervention up to 7.4 years after the randomised controlled trial was started.

Thus, within the primary health care paradigm, where increasing emphasis is placed on nutrition in health promotion and disease prevention, targeting children at a stage when food acceptance patterns are being developed and before lifelong eating habits have become ingrained,[16] appears sensible. This position is supported by evidence of tracking of nutrient intake,[16] obesity [21] and hypercholesterolaemia [22] into adulthood. All of these, plus smoking and hyperglycaemia, are closely related to fatty streaks and the development of atherosclerotic lesions in the second decade of life.[23] More than ten years ago a WHO Expert Committee published a report "Prevention in childhood and youth of adult cardiovascular diseases: time for action", in which the above is acknowledged and the potential for primary prevention programs is outlined.[24]

It follows that the assessment of dietary intakes of children is important for nutrition monitoring, research and intervention efforts.[25] This is mainly the focus of interest in nutritional epidemiology where the relationship between dietary exposure and disease outcome is being studied. Whilst nutrition epidemiologists and community dietitians investigate diets of groups of people, clinical dietitians see dietary assessment as an essential part of the evaluation of the nutritional status of their individual patients,[26] since this forms the starting point of the nutrition care process which consists of assessment, planning, implementation and evaluation.[10]

Measuring diets poses many challenges relating to accuracy and precision. Random and / or systematic errors may occur, the direction and extent of which may vary with the method used and the population and nutrients studied.[27] These methodological limitations lead Beaton [28] and others to conclude: "All dietary assessment methods are imperfect".

Dietary studies in children have even more difficulties because of children's cognitive abilities to record or remember their intake [29] as well as their restricted knowledge of food and food preparation.[30] All dietary measurements should thus be scrutinised for (comparative) validity, including reliability, before general implementation. The Dietary Assessment Calibration / Validation (DACV) Register [31] specifically aims to continually inform and update the international nutrition community of all validation / calibration studies and publications. It was the result of a strong appeal for such research at the First International Conference of Dietary Assessment in Minneapolis in 1993.[31] McPherson et al [25] recently compiled a comprehensive review on validity and reliability studies specifically among school-aged children.

Since full dietary assessments of usual intake of individuals are time-consuming for the participant / client and the researcher / dietitian and therefore costly, there has been a recent interest in short dietary assessments [32] and dietary screeners in the primary care setting.[33, 34] Such tools can be of a general nature, aimed at identifying nutritional risk, usually for undernutrition, [35, 36] or they can be targeted at specific dietary components, for example fat or folate intake. Analysis can be food-based [37, 38] or on the nutrient level of intake. The 'Healthy Eating Index' [39] and the 'Diet Quality Index Revised' [40] are examples of summary measures of overall diet quality, in addition to those reviewed by Kant.[41] Finally, tools can be designed for the population at large or for specific target groups (defined by age, culture, literacy, setting such as hospitals, health condition et cetera) with different time frames (for example recent intake versus usual habits). Validity of a tool depends on its aim,[42] and consequently a dietary screener for usual consumption of a high fat diet by South African children has to be locally developed and validated.

No such tool is currently available; to the contrary, no dietary screeners or short assessment tools, validated for South Africa, were included in the DACV data base up to 9 April 2001, nor was there any mention of validation studies of dietary assessment focusing on fat intake.

In conclusion: A validated dietary screener for fat intake of South African children will fill a research, public health and clinical practice need and may contribute to (cost) effectively managing and preventing the rising prevalence of CNCD's.

1.2 PRACTICAL CONTEXT

The current research project is the first within a research area "Nutritional assessment" established in the Division Human Nutrition, University of Pretoria. The outline of this research area is given in Figure 1.1.

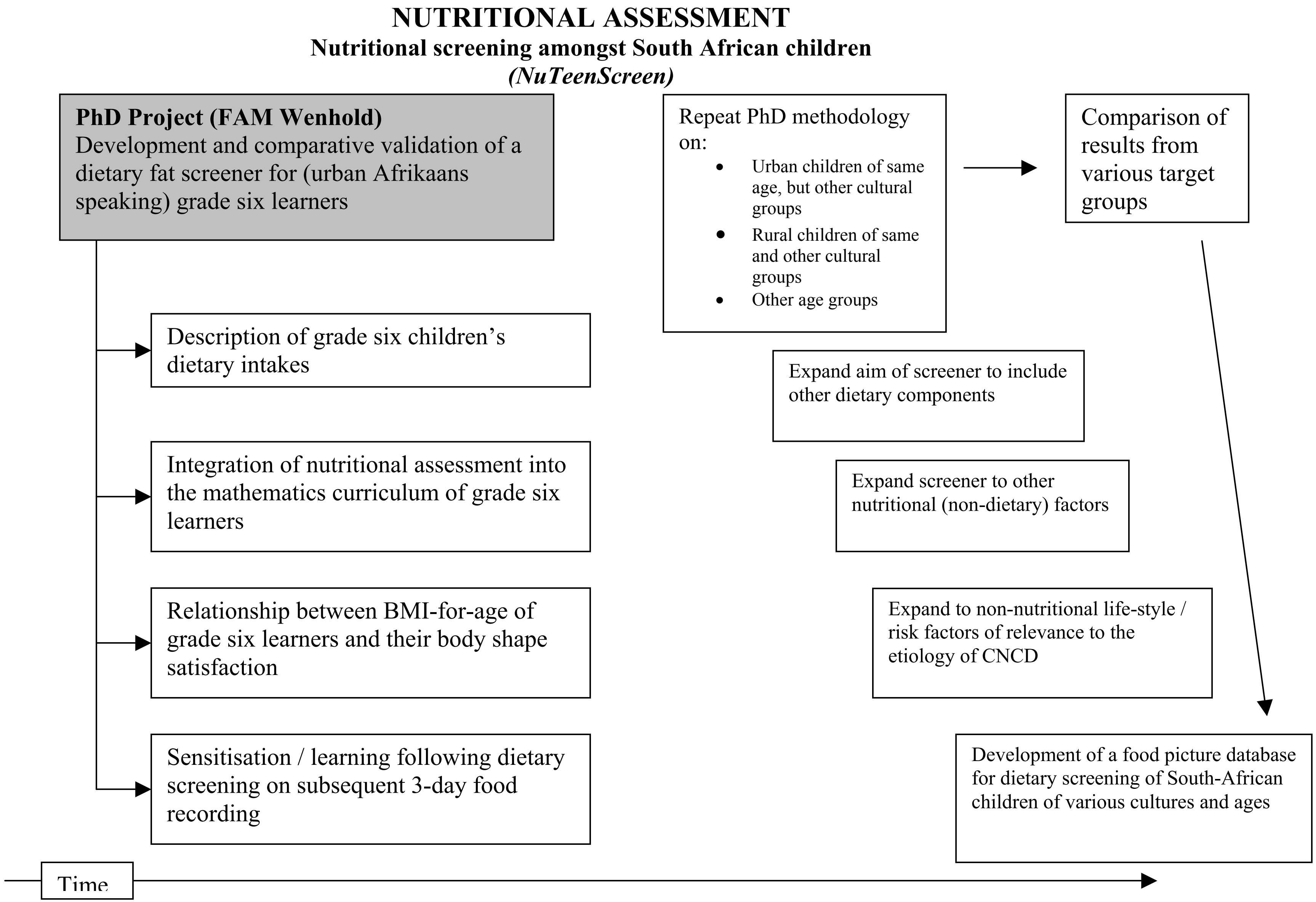


FIGURE 1.1: CURRENT STUDY (shaded box) WITHIN BROADER CONTEXT

1.3 RESEARCH PROBLEM AND SUB-PROBLEMS

The research problem that formed the basis for this study was:

What is the comparative validity of a dietary fat screener (the test method) in grade six learners?

The following sub-problems were formulated:

- What is the reliability of the test method?
  - What is the internal consistency of the test method?
  - What is the test-retest reproducibility of the test method in terms of
    - portion size estimation of all food categories
    - frequency of intake of all food categories
    - category scores of all food categories
    - final scores
    - screener classification?

- What is the validity of the test method compared to a three-day food record?
  - What is the validity of the test method relative to mean daily dietary percent fat energy (PFE) as determined by a three-day food record?
  - What is the validity of the test method relative to mean daily dietary percent saturated fatty acid (PSFE) as determined by a three-day food record?
  - What is the validity of the test method relative to mean daily dietary cholesterol intake as determined by a three-day food record?

- What is validity of the test method compared to the screener as completed by the parents?
  - What is the validity of the test method compared to parental completion of the screener in terms of
    - portion size estimation of all food categories
    - frequency of intake of all food categories
    - category scores of all food categories
    - final scores
    - screener classification?

- What is the validity of the test method when compared to the three-day food record and the screener completed by parents simultaneously?

An overview of the current project is graphically presented in Figure 1.2, which at the same time is intended to clarify the above-mentioned research sub-problems and outline the conceptual framework of the project. The shaded areas represent the sub-problems, whilst the unshaded boxes refer to the developmental evaluation sub-studies and the quality control measures that were performed in this study.

Developmental evaluation sub-studies

TEST METHOD (Test and re-test)

Food categories	Portion size score (1,2 or 3)		Frequency score (0,3 or 7)		Category score (0-21)
Meat		x		=	
Eggs		x		=	
Dairy, **m**ilk		x		=	
Dairy, **c**heese		x		=	
Dairy, **d**essert		x		=	
Fried foods		x		=	
In baked goods		x		=	
Convenience foods		x		=	
Table fats		x		=	
Snacks		x		=	
Final score (0-210)					
Classification: 'High fat' or 'Prudent'					

Anthropometric and design quality control

Reference method 1:

FOOD RECORD

PFE PSFE Cholesterol

Classification: 'High fat' or 'Prudent'

Developmental evaluation sub-study

Reference method 2: SCREENER BY PARENTS

Food categories	Portion size score (1,2 or 3)		Frequency score (0,3 or 7)		Category score (0-21)
Meat		x		=	
Eggs		x		=	
Dairy, **m**ilk		x		=	
Dairy, **c**heese		x		=	
Dairy, **d**essert		x		=	
Fried foods		x		=	
In baked goods		x		=	
Convenience foods		x		=	
Table fats		x		=	
Snacks		x		=	
Final score (0-210)					
Classification: 'High fat' or 'Prudent'					

FIGURE 1.2 CONCEPTUAL FRAMEWORK

1.4 TERMINOLOGY

In Table 1.1 the core concepts and abbreviations as used in this study are (operationally) defined. Where no strict differentiation between terms was applied, stipulating the synonyms used indicated this.

TABLE 1.1: TERMINOLOGY, (OPERATIONAL) DEFINITIONS AND ABBREVIATIONS FOR THIS STUDY
(Arranged alphabetically; Concepts in italics are cross-referenced)

Terminology / abbbreviation	Description and / or operational definition
ADA	American Dietetic Association
AMDR	Acceptable macronutrient distribution ranges
Agreement	For categorical data: A *match* between corresponding variables in two methods / administrations, for example portion size in the test method and in the reference method, or *classification* in the first and the second administration of the screener; expressed in terms of percentage identical responses and chance-corrected kappa statistic For continuous data: Inferred from statistical results, for example (linear) associations (comparisons of rank orders), comparisons of means and assessing the extent of differences, including indications of random and systematic errors, and the Bland-Altman method
Anthropometry	Study of the size and dimensions of the human body In this study: Weight and height of the children, combined into the following indices: • Weight for age • Height for age • Body mass index for age Interpreted by expressing in terms of mean percentiles and Z-scores with the CDC 2000 growth charts as reference
Assessment	The numerical value given to some physical property (for example weight) or behaviour (for example dietary intake) Synonyms: Measurement, estimation, prediction, evaluation, determination
Bias	For categorical data: The absence of symmetry in a cross tabulation of corresponding variables as indicated by the McNemar test for symmetry For continuous data: The difference between values obtained in the test method and the reference method [43] or between two administrations of the test method Synonym: Systematic error (see text in review of literature) In this study: Dependent on the phase in the validation process, for example in the test-retest reproducibility study this could refer to portion size scores in first versus second administration (categorical data), or in the child-parent comparative validation this could refer to final scores of children minus the final scores of parents (continuous data)

Terminology / abbbreviation	**Description and / or operational definition**
Categorised weekly consumption	Conversion of reported times per day or week consumption *(frequency of intake)* in the test method and reference method 2 (screener by parents) into the following categories: • Less than once per week • Once or more (up to three times) per week • More than three times per week
Category score	In the test method and the screener completed by the parents: The mathematical product of the *portion size score* and the *frequency of intake score* of each food category
Chronic non-communicable diseases (CNCD)	Chronic (in contrast to acute) non-contagious diseases, also called 'diseases of lifestyle', for example cardiovascular disease, including hypertension, type 2 diabetes mellitus, certain cancers (for example colon, breast, prostate)
Classification	Dichotomisation of *final scores* in test method and screener by parents (reference method 2), and of *PFE*, *PSFE* and cholesterol intake (reference method 1) to *'high fat' or 'prudent'*
Comparative validation	The relation between a less detailed method of dietary assessment to a more detailed method, assumed or shown to more closely reflect the truth Synonyms: Calibration, relative validation, standardisation, congruent validation <u>In this study</u>: Reported intakes from the test method relative to the chosen reference methods (three-day food record and screener by parents), where the reference method(s) were assumed to reflect true usual fat intakes (that is the truth)
Construct	The unobservable (or latent) trait being measured by the questionnaire. The construct or trait is measured along a continuous scale. Synonyms: Trait, domain, latent variable, theta, characteristic, attribute <u>In this study</u>: Usual fat intake
Developmental evaluation sub-study	First stage evaluation of a test method where the adequacy of a tool as such is assessed prior to field testing and comparative validation in the target population[44] <u>In this study</u>: The sub-studies described in the chapter 'Development and developmental evaluation' that is: • Sub-study 1: Content and face validity (Test method) • Sub-study 2: Reference portion size (Test method) • Sub-study 3: Portion size estimation aids (Test method) • Sub-study 4: Frequency of intake (Test method) • Sub-study 5: Food record (Reference method 1)
Dietary fat screener	Short method for *assessing* fat intake (see *screener*)
DRI	Dietary Reference Intakes
Final score	Mathematical sum of the ten *category score*s in the test method and in reference method 2 (screener by parents), in both cases potentially ranging from 0 to 210

Terminology / abbbreviation	**Description and / or operational definition**
Food category	Line item in the food frequency type test method and reference method 2 (screener by parents), referring to **M**eat, **E**ggs, **D**airy (milk), **D**airy (cheese), **D**airy (dessert), **F**ried foods, fats **In** baked goods, **C**onvenience foods, **T**able fats
Food consumption	Food and drink ingested by participants Synonyms: Food intake; dietary intake
Food frequency questionnaire (FFQ)	A list-based interview procedure during which the participant recalls how often specified foods or food groups are eaten per day, per week or per month. It may include quantitative assessment of usual portion size (and is then more accurately called a 'food frequency and amount questionnaire') <u>In this study</u>: A ten item (food category) list with per day or per week frequency of intake format, and relative portion size reporting
Food record	A written record / diary of current food and drink intake by the participant concurrently or immediately following the eating occasion for the specified *recording period*
Frequency of intake	Number of times a food category is usually consumed per day or per week
Frequency score	Point score given to *categorised weekly consumption* in the test method and reference method 2 (screener by parents) • Less than once: 0 points • Once or more (up to three times per week): 3 points • More than three times: 7 points
Grade six child	Learner (scholar, pupil, student) in the sixth grade (Intermediate Phase of the South African Department of Education C2005 for schools); typically 12 years old
High fat / prudent diet	<u>Test method and screener by parents</u>: If the *final score* obtained exceeded 68, it was classified as 'high fat'; conversely it was classified as 'prudent' <u>Reference method 1</u> (Three-day record): If the following conditions were met based on the mean daily intake in the three-day *food record*, the diets were classified as 'high fat'; conversely the diets were classified as 'prudent': • Percentage total fat energy (PFE) > 30 • Percentage saturated fat energy (PSFE) > 10 • Cholesterol intake => 300mg
Inter-individual	Between persons / participants
Internal consistency	Homogeneity, uni-dimensionality of a *scale* <u>In this study:</u> Internal consistency together with reproducibility were taken as indicators of reliability of the test method. Internal consistency was measured by item total correlations, Cronbach's coefficient alpha and the split half method (equivalent forms approach)

Terminology / abbbreviation	**Description and / or operational definition**
Intra-individual	Within the person / participant
IOM	Institute of Medicine
Item	Question in a *scale* In this study: The ten food categories (**M**eat, **E**ggs, **D**airy [milk, cheese and dessert], **F**ried foods, fats **I**n baked goods, **C**onvenience foods, **T**able fats, **S**nacks) contained in the dietary fat screener, being the line items in a FFQ
Match	When categorical data (for example reported portion size, frequency of intake or fat intake classification) were the same in two or more assessments (for example test and re-test or child and parent) Expressed as percentage identical responses or 'perfect agreement' in tables, or as overlap areas in figures
MEDFICTS	Dietary assessment tool developed by National Cholesterol Education Program (NCEP) Acronym for food categories contained in test method, that is • **M**eat • **E**ggs • **D**airy (milk) • **D**airy (cheese) • **D**airy (dessert) • **F**ried foods • fats **I**n baked goods • **C**onvenience foods • **T**able fats
NCEP	National Cholesterol Education Program
Negative predictive value (NPV)	The probability of the person not having the condition when the test is negative[45]. The formula for negative predictive value is TN / (TN + FN) where TN and FN are the number of true negative and false negative results respectively In this study: Proportion of individuals who truly consumed a prudent diet according to the three-day food record, out of all who tested negative with the test method
Odds ratio (OR)	The ratio of the odds of exposure among the cases to the odds in favour of exposure among the controls [45]
Overall predictive value (OPV)	The proportion of predictions that are true positives and negatives In this study: Proportion of individuals that truly consumed a high fat and prudent diet
PFE (Percentage fat energy)	Mean daily total dietary fat intake (in grams) from the three-day records was converted to an energy (kJ) equivalent by multiplication by 37.8. PFE was then calculated by expressing mean total fat energy as a percentage of mean daily energy intakes.
Physical Activity Level (PAL)	Ratio of dietary energy intake to basic metabolic rate (BMR); Part of the 'Goldberg' cut-off for performing quality control in self-reported energy intake [46]

Terminology / abbbreviation	**Description and / or operational definition**
Portion / Serving size	In the test method and reference method 2 of this study: • A small portion was defined as half as much or less than the *reference portion* • A medium portion was defined as about the same amount than the *reference portion* • A large portion was equivalent to one-and-a-half times or more the size of the *reference portion*
Portion size estimation aid (PSEA)	Two-dimensional (2D) and three-dimensional (3D) props to help participants gauge intake quantities consisting of geometrical shapes, photos and household measures (measuring cups and spoons) <u>In this study</u>: Full list of PSEA used are described in text
Portion size score	Point score given to reported *portion size*: • Small = 1 point • Medium = 2 points • Large = 3 point
Portion, serving	The amount of food that a person reports as being the quantity usually consumed. There is no standard portion size and no single right or wrong portion size <u>In this study</u>: No official standardisation is available for South Africa and in the Afrikaans language no differentiating terminology is generally accepted. Thus portion and serving were used interchangeably (and always explained), except where to specific studies was referred
Positive predictive value (PPV)	The positive predictive value of a test is the probability of the person having the condition when the test is positive.[45] The formula for positive predictive value is TP / (TP + FP) where TP and FP are the number of true positive and false positive results respectively. <u>In this study</u>: The likelihood that individuals categorised by the test method as having a high fat diet had a diet that was high in fats according to the 3-day record
Primary / elementary school	Typically school in the South African Department of Education school system accommodating grades 1 to 7
Percentage fat energy (PFE)	Mean daily, total fat intake (in grams) from the three-day records was converted to an energy (kJ) equivalent by multiplication by 37.8. PFE was then calculated by expressing total fat energy as a percentage of mean daily energy intakes
Percentage saturated fat energy (PSFE)	Mean daily, saturated fatty acid intake (in grams) from the three-day records was converted to an energy (kJ) equivalent by multiplication by 37.8. PSFE was then calculated by expressing saturated fatty acid energy as a percentage of mean daily energy intakes
RDA	Recommended Dietary Allowance
Recording period	One of three specific sets of consecutive days during which children kept *food record,* that is either one of the following: • Thursday, Friday and Saturday • Tuesday, Wednesday and Thursday • Sunday, Monday and Tuesday

Terminology / abbbreviation	Description and / or operational definition
Reference method	Comparison (more detailed or accurate) method assumed to be superior to the test method Synonyms in literature: (gold) standard; 'outcome for that which is to be predicted or detected by the screener' [47] In this study: Reference method 1: The three-day food record Reference method 2: Screener completed by parents
Reference period	The time span in relation to which intakes in a FFQ are reported In this study: From the beginning of the academic year (January) to the time of assessment (that is September), representing nine months and assumed to reflect usual intake of fat
Reference portion	The amount given to participants relative to which the own intake had to be reported (see *portion size*)
Relative risk (RR)	The ratio of the risk occurrence of a condition among exposed people to that among the unexposed [45] Synonym: Risk ratio
Reliability	Synonyms: Absence of random error, precision, consistency, reproducibility, repeatability, dependability; see text (literature study) for more detail In this study: Conceptualised in terms of *internal consistency* and test-retest *reproducibility*
Reproducibility	The extent to which a method produces the same results when applied repeatedly in the same situation Synonym: Repeatability In this study: Test-retest reproducibility of the test method was measured and, together with (statistically quantified) *internal consistency*, were taken as indicators of *reliability*
Scale	Synonyms: Measure, questionnaire, instrument or test; but also apparatus or equipment In this study: • The dietary fat screener was taken to be a scale, which measured a single *construct* or domain, namely fat intake • The term 'scale' was also used to refer to the instrument / physical equipment used for weighing foods (Soehnle scales) and to measure the children's body mass (Tanita scale)
Screener	Synonyms: Short assessment, 'low intensity method',[48] simple indices as alternatives to more complex methods, [49] 'short-cut method' [50] In this study: *Dietary fat screener* (the test method)
Sensitivity	The sensitivity of a test is the probability of a positive test in people with the condition[45]. The formula for sensitivity is TP / (TP + FN) where TP and FN are the number of true positive and false negative results respectively. Sometimes called the "true positive fraction" and calculated as Prob [screener positive | outcome positive] [47] In this study: Proportion of individuals with high fat intake in the three-day food record who were correctly classified by the test method as having a high fat intake

Terminology / abbbreviation	Description and / or operational definition
Specificity	The specificity of a test is the probability of a negative test in people without the condition.[45] The formula for specificity is TN / (TN + FP) where TN and FP are the number of true negative and false positive results respectively. Sometimes defined within the context of conditional probability, where the "false positive fraction" = 1- specificity = Prob [screener positive | outcome negative] [47] In this study: Proportion of individuals following a prudent diet according to the three-day food record correctly classified by the test method as not being at risk of high fat intake
Test method	The new or simpler method or method of unknown performance Synonyms: Pepe et al [47] used the generic term 'marker' for a factor, score or biomarker used for screening In this study: The dietary fat screener
Tool	The means by which dietary assessment and screening were performed Synonyms: Instrument, questionnaire, test, measure In this study: Examples of tools are the dietary fat screener or the three-day food record
Uni-dimensionality	The set of questions are measuring a single continuous latent variable (*construct*) In this study: All food categories (items) in the test method measure fat intake
WHO	World Health Organisation

CHAPTER 2: REVIEW OF LITERATURE

Chapter overview

2.1 NUTRITION OF PRIMARY SCHOOL CHILDREN
2.1.1 Eating habits of children
2.1.2 Dietary fats in childhood nutrition

2.2 DIETARY ASSESSMENT OF CHILDREN
2.2.1 Overview
2.2.2 Cognitive abilities of children affecting dietary assessment
2.2.3 Integrating nutrition and dietary assessment into the school environment
2.2.4 The food frequency questionnaire (FFQ) as basic format of the test method
2.2.4.1 Description
2.2.4.2 Aims
2.2.4.3 Strengths and limitations
2.2.4.4 Development of FFQ's
2.2.4.4.1 Item list
2.2.4.4.2 Quantification

- Reference portion size and portion size estimation aids (PSEA)
- Frequency of intake
- Nutrient content of items

2.2.4.5 FFQ for children
2.2.4.6 FFQ for fat intake
2.2.5 The food record
2.2.5.1 Description
2.2.5.2 Strengths and limitations
2.2.5.3 The food record as reference method
2.2.5.4 The food record for children

2.3 NUTRITIONAL AND DIETARY SCREENING
2.3.1 Definition and characteristics
2.3.2 Aims
2.3.3 Examples of screeners
2.3.4 Fat screeners

2.4 RELIABILITY AND VALIDITY IN DIETARY ASSESSMENT AND SCREENING
2.4.1 Variability and error in dietary assessment
2.4.1.1 True variability
2.4.1.2 Error
2.4.2 Reliability
2.4.2.1 Definition
2.4.2.2 Types of reliability
2.4.2.3 Measurement of reliability
2.4.3 Validity
2.4.3.1 Definition and principles
2.4.3.2 Types of validity
2.4.3.3 Validity of (dietary) screening tools
2.4.4 Validation studies
2.4.4.1 Background
2.4.4.2 Validation studies in children
2.4.5 Factors influencing validity and reliability
2.4.6 Implications

2.1 NUTRITION OF PRIMARY SCHOOL CHILDREN

There is general consensus that appropriate eating habits of children are important for optimal physical and cognitive development, the attainment of healthy weight and the reduction of the risk of CNCD.[16] In order to achieve nutritional health many organisations and countries have set up dietary guidelines, including South Africa, where country-specific, evidence-based food-based guidelines for people seven years and older were officially approved and adopted by government in 2003.[14]

In 2002 the Institute of Medicine's (IOM) Food and Nutrition Board released the Dietary Reference Intakes (DRI) for energy, carbohydrates, fat, fatty acids, and cholesterol, thereby updating the Recommended Dietary Allowances (RDA). Acceptable macronutrient distribution ranges (AMDR) as a percent of energy intake for carbohydrate, fat, and protein for children are as follows: [10, 16]

- Carbohydrate: 45% to 65% of total energy
- Fat: 30% to 40% of energy for 1 to 3 years
  25% to 35% of energy for 4 to 19 years
- Protein: 5%- 20% for young children
  10% to 30% for older children

Knowledge of current eating habits of children relative to these dietary guidelines and reference intakes is an important starting point for appropriate intervention for maintenance or improvement of the nutritional status of children. In the following section some major findings of relevance to the current research context are presented, first from the international literature and then from South Africa.

2.1.1 Eating habits of children

Dietary intake data of children on the international arena suggests the following:

Amongst eleven to 18-year-old United States (US) adolescents from 1965 to 1996 total energy intakes decreased, as did the proportion of energy from fat (from 39% to 32%) and saturated fat (15% to 12%). There were concurrent increases in consumption of higher fat potato and mixed dishes (pizza and macaroni cheese), lower fat milk replaced higher fat milks, but total milk consumption decreased by 36%.[51] Ten-year results of the NHLBI Growth and Health Study showed that for girls total and saturated fat and cholesterol intakes had decreased with age. In all cases the decrease was more in white girls than in black girls. A substantial percentage of both ethnic groups had not yet reached NCEP goals in terms of PFE, PSFE and cholesterol.[52] Even children known to be at high risk for cardiovascular disease (based on the NCEP criteria) were no more likely to meet guidelines for heart-healthy diets than were children at low risk.[53]

Dwyer et al [54] investigated the eating patterns of US adolescents and found that an increase in eating occasions was common. It was associated with increased energy intake but a reduced relative amount of total and saturated fat consumed. 'Grazing' may be the modal behaviour of children who also increasingly make their own food decisions.[55] A high percentage of daily food intakes of children occurs at schools. In the US, school stores were found to sell primarily snacks with high fat and sugar content.[56] Amongst participants in the Bogalusa Heart Study, Nicklas et al [57] reported a striking change in meal patterns over a 21-year period: They observed increases in the number of meals eaten away from home and at restaurants, decreases in home dinners, snacking and total eating episodes.

A recent review of current dietary trends and quality, including evidence of tracking of nutrient intake in children, as well as meal patterns, frequency and portion size information in US children aged two to eleven years has been compiled by the Ameican Dietetic Association (ADA) [16] and may also be relevant for older children and adolescents.

Hackett et al [58] studied eating habits of eleven and twelve-year-old children before and after the start of a healthy eating campaign in the United Kingdom (UK). They found favourable changes, but also that the "case for encouraging changes in the eating habits of children is compelling". This was confirmed by national and regional surveys of 11 to 14-year old UK children where the most popular items emerged were the least desirable foods: Confectionary (crisps, chocolates, sweets), biscuits and cakes, chips and sugar-flavoured fizzy drinks.[59] The dietary trends amongst Scottish school children in the 1990s suggest increased intakes of fruits and vegetables (but still below recommendations) and concomitant increases in high-fat and high-sugar foods, the latter particularly amongst boys and children from lower socio-economic groups.[13] A study of 158 German primary school children showed that they consumed 42% of energy from fat with about 50% as saturated fat.[60]

For South Africa national data for children are limited to the age group one to nine years.[61] Very few studies addressing children are included in a report on South African food consumption studies undertaken amongst different population groups between 1983 and 2000.[62] The THUSA Bana study focused on people in transition in the North West Province and included 1257 children, of which 868 in the age group ten to 13. Maize porridge, white sugar, brown bread, full cream milk and white bread were the most commonly consumed foods.[62] Over 40 years ago (1962) the nutritional status of six to eleven year-old white primary school children in Pretoria was surveyed in depth. At that stage percentage energy as fat, protein and carbohydrate were reported to be 35%, 12% and 53% respectively. It was concluded that the nutritional status was equivalent to that of an "affluent population group." [63]

2.1.2 Dietary fats in childhood nutrition

One of the US Dietary Guidelines 2000 for the general population states: "Choose a diet that is low in saturated fat and cholesterol and moderate in total fat." [12] The equivalent in the food-based dietary guidelines for South Africa is: "Eat fats sparingly." [13] The NCEP report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents [17] recommends the following intakes in the Step 1 diet:

- Total fat — Average of no more than 30% of total energy
- Saturated fatty acids — Less than 10% of total energy
- Polyunsaturated fatty acids — Up to 10% of total energy
- Monounsaturated fatty acids — Remaining total fat energy
- Cholesterol — Less than 300mg/d
- Carbohydrates — About 55% of total energy
- Protein — About 15-20% of total energy
- Energy — To promote normal growth and development and to reach or maintain desirable body weight

The American Heart Association Guidelines [15] reiterated the above as being population guidelines, which should also apply to children and adolescents.

Butte [64] reviewed the optimal fat intake for children against the background of their energy requirements. They state that the current recommendations of 30% of energy from dietary fat for children older than two years are sufficient for adequate growth. Lower intakes may be associated with micronutrient inadequacies. Higher intakes may lead to increased energy intakes and increases in body fat, but conflicting data are available.

In industrialised countries the conclusion thus seems to be that the primary prevention of CNCD should begin in childhood, [15, 16, 65] but there is no agreement on the most appropriate application, for example the US versus Canada regarding the best age from which to recommend these intakes.[66] In many developing countries too low fat intakes may be a greater concern,[67] leading several researchers [68, 69, 70] to point out the possible dangers of dietary fat restriction for children. Nevertheless, in a contra-point Lytle [71] defended a low-fat diet for healthy children and Van Horn [72] also reconfirmed the NCEP stand.

The Institute of Medicine in the most recent release of Dietary Reference Intake values seems to have accommodated both sides of the coin by formulating 'Acceptable Macronutrient Distribution Ranges' (AMDR), defined as 'a range of intakes for a particular energy source that is associated with reduced risk of chronic disease while providing adequate intakes of essential nutrients'. [10]

To address the nutritional excesses and deficiencies of South Africa, a land of contrasts, Vorster et al [73] have also suggested that more specific guidelines be adopted, for example instead of advising that less than 30% of energy should come from fat, ranges, for example between 25% and 30%, or to aim for 30% should be considered.

2.2 DIETARY ASSESSMENT OF CHILDREN

2.2.1 Overview

Against the backdrop of evidence that many of the risk factors for the development of CNCD, for example nutrient intake,[74] obesity [21] and hypercholesterolaemia,[75] track from childhood into adulthood, dietary assessment of children is important in nutrition monitoring, research, and in clinical and community-based interventions.

Stang [76] has compiled a practical overview of assessment of nutritional status in clinical practice, including dietary assessment, of adolescents. In the research context methods that can be used for dietary assessment range from very sophisticated individual-level investigations suitable for metabolic wards to 'bird's eye views' aimed at describing diet on group level. Bingham [77] has published a comprehensive review of the dietary assessment methods for use on individuals. Several overviews [30, 78, 79] focused their discussion on dietary assessment of children.

Apart from the general accuracy and precision issues of dietary evaluation,[80] studies in children pose additional challenges.[79, 81] Three recent reviews emphasised the importance of establishing reliability and validity of dietary assessments specifically in children.[25, 30, 81]

The *duplicate portion technique* has sometimes been described as giving very accurate information on the nutrient level. Isaksson [82] reviewed the principles involved. When total dietary fat and fatty acid intake measured by chemical analysis of duplicate diets were compared to nutritional database analysis of estimated dietary records, collected over the same three-day period, lack of agreement was found.[83]

Unobtrusive *observation* in assessment of children's dietary practices minimises self-report problems and has been considered as a 'gold standard' against which other measures of behaviour could be compared. The use thereof has been described by Baranowski and Simons-Morton.[79]

The four methods most commonly used for assessing diets of individuals are the food record, the 24-hour recall, the food frequency questionnaire (FFQ) and the diet history. The first two methods describe current intake and are meal-based, whereas the latter two describe past or usual diet. All of these have been used in assessment of diets of children.[25, 78] One of the major long-term studies involving children showed the feasibility of implementing a variety of dietary assessment methods among pre-adolescent children without relying primarily on parental reports, [84] but for younger children parents are usually included either as surrogates or in addition to the child report.

New approaches, for example using the computer, telephones and tape recorders to record children's food intake are being investigated.[85, 86] Diet analysis tools are increasingly available online.[87]

The FFQ and food record are discussed in more detail in later sections of this literature review.

2.2.2 Cognitive abilities of children affecting dietary assessment

Self-report of diet necessarily involves cognitive processes, although for many years limited research has focused on either adults' [88, 89] or children's [90] cognitions in regard to food. Recently various research groups have started paying attention to this aspect.

It can be assumed that children in the four major periods of cognitive development (that is sensorimotor [birth to two years], pre-operational [two to seven years], concrete operation [seven to eleven years] and formal operations [eleven years and beyond]) differ in the way they process food information. The latter age is about the age from which children have been shown to provide reasonably accurate dietary information.[25, 78, 81, 90]

In order to systematically analyse the mental activities involved in food recall, Baranowski and Domel [90] have proposed a model of a child's cognitive processing of food information (Figure 2.1). This model is the result of combining cognitive psychology with survey methodology in order to optimise the collection of valid food intake data. In the model the recall of foods and the number of portions of such foods is addressed, primarily on the short-term. The cognitive skills involved in recalling frequency of intake (for example event equalisation, estimation of frequency, averaging) are not explicitly covered, but can be inferred. The model analogises human cognition to methods by which a computer processes, categorises, stores, and retrieves information. Apart from the implications for developing new dietary assessment tools, this model provides a starting point for categorising errors that can be encountered with children's self-report of diet. As is evident from Figure 2.1, these errors can be related to attention, perception (or interpretation), organisation, retention, retrieval, and response (printed in italics in the Figure 2.1).

Noticing, that is paying *attention* to food eaten is a prerequisite for future recall. Inattentiveness may result in underreporting. The model also shows that paying attention to the request for dietary information is the critical starting point for valid recall. By increasing the interest in the task of remembering what was eaten, the quality of response can be improved.[90] Question comprehension is a critical cognitive stage in any dietary assessment involving recall. [91] In a FFQ it is essential that the participants understand the question and know how to report consumption frequency, portion sizes and compute average yearly use of seasonal items.[91]

The foods consumed must be *perceived* by the child to be the same thing that the researcher meant. In their study involving fifth to seventh graders, Koehler et al [92] listed food knowledge, preparation and vocabulary as instrument-related factors influencing the validity of a dietary assessment tool. In this respect, the use of pictures has been shown to reduce misunderstanding. Picture-to-picture matching appears to be superior to picture-to-word matching, and pictures appear to trigger memory, where words have not.[88]

Organisation, in the dietary recall context, refers to the grouping of foods in long-term memory. It may be that children classify foods differently to adults, for example by using functional criteria (that is meals versus snacks), nutritional or healthful criteria, or sweetness [90] instead of, for example, the basic food groups. Furthermore, different children may organise foods differently, use different reasons for categorisations, and this may be affected by developmental stage. Baranowski et al [90] and Koehler et al [92] reported that some children had difficulty understanding the wording of food categories on a food frequency form, for example when deciding in which category particular foods should be placed. Based on this rationale, some researchers (for example Kohlmeier [88]) have started rearranging their FFQ to a meal-based, rather than a list-based format.

Retention refers to memory and is related to time lapsed between the actual consumption and the request to recall the intake. In the case of children it was found that food memory decay varied by food group. However, underreporting appears to be more common than overreporting, even though this may, in part, be the result of researchers failing to differentiate between 'underreporting' (reporting one half of a banana when a whole banana was eaten) and 'failure to report' (reporting that no snack was eaten when in fact a banana was eaten).[90] Baxter and Thompson [93] found that the cognitive burden of recalling items eaten at school lunch as part of a 24-h recall was greater than that of recalling school lunch items as single meal. Thus the latter yielded more accurate information. Even under the best conditions (for example reporting within 90 minutes after the meal) children have difficulties reporting their intakes.[94] This is, however, not unequivocally accepted: In a longitudinal study Dwyer and Coleman [95] found that there was not necessarily a clear decline in accuracy of report over time when the same subjects were studied over four decades. Nevertheless, the distorting effect of current diet on recall of past food consumption was revealed.

The process of *retrieval* involves obtaining information out of long-term memory into short-term memory to form a response. In this stage interference can be a problem.[91] As the time interval is increased over which diet must be recalled, interference will also be increased. It is unlikely that food information is stored separately, but it is probably embedded in other events.[90] This is why the use of event prompts (for example a party, sports event) can improve a 24-h recall. Domel [29] explored how children remembered food intake and identified several memory retrieval-response categories, including visual imagery, usual practice, behaviour chaining, preference, food labels and so forth. The effect of different types of prompting on the accuracy of children's food recalls has consequently been investigated.[96] These researchers found that among first grade students, specific prompting in terms of preference, food category or visual cues resulted in more harm that good. Among fourth-grade learners prompting for food categories resulted in some improvement in accuracy. When children report eating standard portion sizes rather than the real amount eaten, this can result in overreporting of low intakes and underreporting of high intakes, the so-called flat slope syndrome.[90] Prompting children to report foods eaten over the previous 24 hours in reverse versus forward order improved omission and intrusion rates of fourth-graders' recalls, particularly for boys, but the overall error rate (omission plus intrusions) remained high.[97] Ensuring recognition of food items (either on a word-based list or on pictures), which is cognitively the core task in food frequency questionnaires,[89] aids the retrieval process.

Finally, *response* refers to the way in which children wish to present themselves to others. Social desirability plays a role here and was found to occur when children underreported candy consumption and over-reported vegetable consumption in a telephone recall compared to their parents' reports.[90] If an event is considered as embarrassing, sensitive in nature, threatening or divergent from the respondent's self-image, it is less likely to be reported.[91]

Thus, theoretically, memory and cognition are required for completing a FFQ, because participants must first recognise the food item. The consumption of each item must then be considered, the information over the reference period (for example one year) be integrated and finally the average frequency of food use computed. All of these processes are interlinked as indicated in Figure 2.1.

In practice Drewnowski [98] has, however, argued that “reality is beside the point: FFQ's reflect a long-established predisposition toward a mental image of a given food.” He thus implies that the cognitive processes are overemphasised and that in actual practice food preferences or attitudes are measured by FFQ and not ‘usual consumption’.

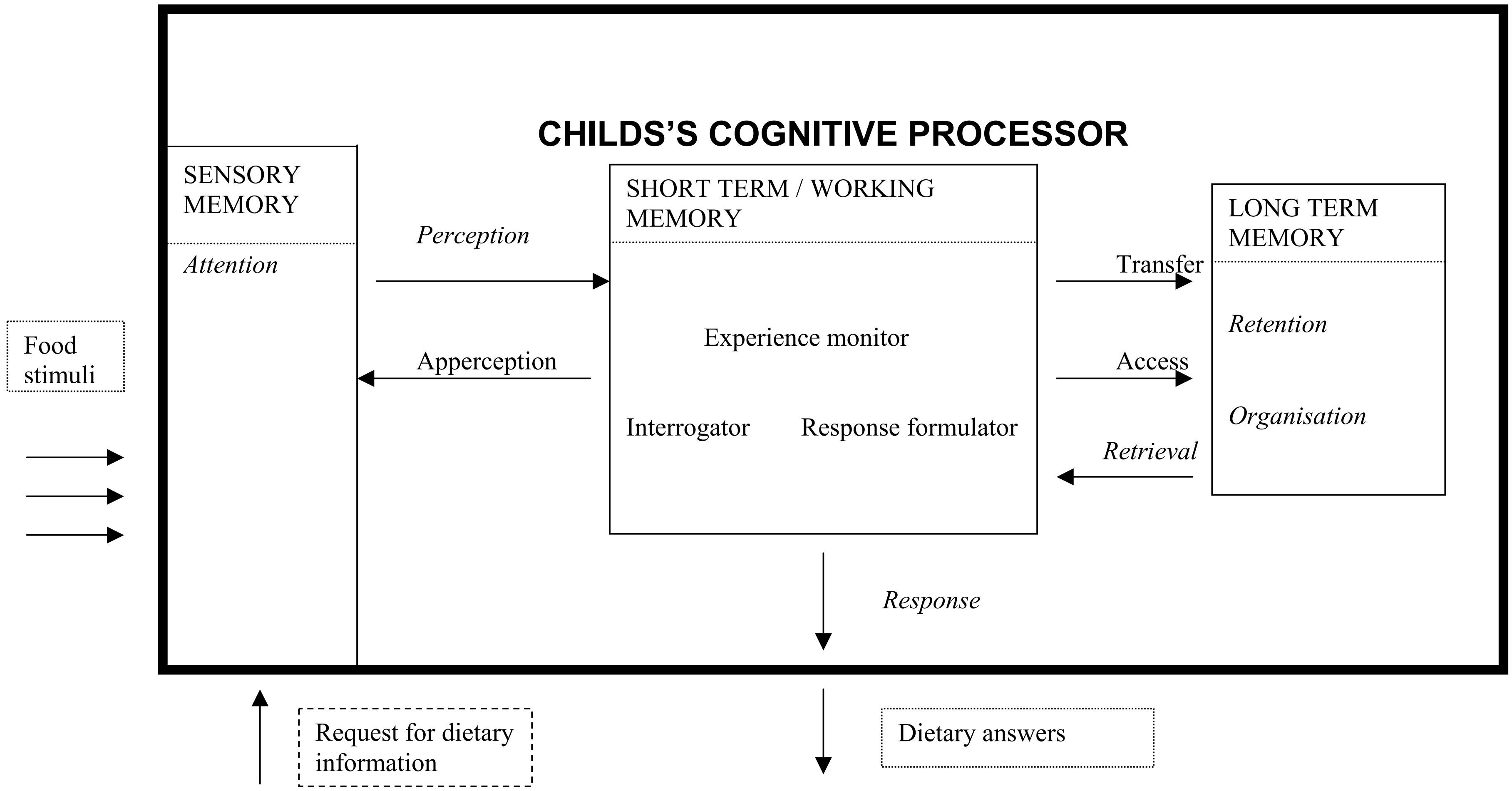


FIGURE 2.1: MODEL OF A CHILD'S COGNITIVE PROCESSING OF FOOD INFORMATION [90]

2.2.3 Integrating nutrition and dietary assessment into the school environment

Frank [55] and Story et al [99] identified multiple environments of children, all of which influence eating patterns and also the choice of method for collection of dietary information. Apart from the personal, the home, the media / entertainment environments, as well as fast food eateries and the food industry, they highlighted the school environment, where most children spend six to eight waking hours per day.

Schools have consequently been identified as ideal settings for promoting health and lifelong healthy eating amongst children for the following reasons: [65]

- Schools can reach almost all children
- Schools can provide opportunities to practice healthy eating
- Schools can teach children how to resist social pressure. Since eating is a socially learned behaviour, social (peer) pressures that discourage healthy eating can be directly addressed and positive peer pressure can be reinforced
- Skilled personnel are available. Teachers can receive nutrition knowledge and then use their instructional skills to reach the children
- Research suggests that school-based nutrition education programmes can improve eating behaviour of children [65]

The use of schools for achieving nutrition aims has repeatedly been documented in the international literature for primary, middle and secondary schools, in urban and rural settings, for CNCD risk reduction [100, 101] and for addressing and preventing obesity [102] and undernutrition, as stand-alone nutrition education projects and as integrated programmes where, for example, nutrition education, food provision and promotion of physical activity are jointly included in a bigger programme.[103] The Health Promoting Schools concept of the WHO is another example of integration.

Florencio [104] reviewed various approaches to school-based feeding programmes around the world. Those, which aim to tie-in with nutrition education range from a focus on protein-energy malnutrition, to micronutrient deficiencies and obesity. Sometimes the emphasis is on knowledge, but values, attitudes and skills have also been the objective. Strategies used include formal, class-room-based and also informal, extra-curricular activities, with varying emphasis on the involvement of the child, the existing teachers, parents and the community.

A review on using the school environment for promoting physical activity and healthy eating was published by Wechsler et al.[105] They concluded that “enough is known from theory, practice and research to suggest that school-based environmental strategies to promote physical activity and healthy eating among young people merit implementation and ongoing refinement”. A practical, theory-based attempt to establish school nutrition advisory councils as an integral part of a school environment approach towards nutrition promotion has been proposed and described by Kubik et al.[106]

Many recent school-based nutrition programmes are placed within the framework of the US Center for Disease Control's recommendations for school health programmes promoting healthy eating. The seven recommendations are policy formulation, curriculum development for nutrition education, instruction for students, integration of food service and nutrition, training of school staff, family and community involvement, and programme evaluation.[65]

In South Africa the national Integrated Nutrition Programme (INP) reflects current policy in this regard. It includes the Primary School Nutrition Programme (PSNP), which aims to provide food supplementation, nutrition education and parasite control to the poorest of the poor, in an attempt to address short term hunger, the high prevalence of inadequate growth and micronutrient deficiencies (referring to iron and vitamin A), and to improve school attendance. Thus cooperation between nutrition (as part of the Department of Health) and the Department of Education is officially encouraged as part of intersectoral collaboration.[107]

Education in South Africa is in the process of moving towards outcome-based education (Curriculum 2005). This approach favours, amongst others, practical application, skills development and real life problem solving as desirable outcomes of education. The aim is to have integrated curricula.[108]

Nutrition education and assessment in the school context are, however, not without challenges: Teachers and administrators often view it as a loss of classroom instruction time.[55] This negative attitude can be reduced if the message and measurements are incorporated into classroom instruction.[55] Numerous possibilities exist for embedding or integrating nutrition into other subjects and various school subjects could be ‘vehicles’ for carrying nutrition content.[104]

Different teaching and learning strategies can be employed for doing this. Among the elements identified as contributing to the effectiveness of nutrition education and nutrition assessment for school-aged children are the following: [92, 104]

- The use of developmentally appropriate learning experiences and materials
- Activity-based teaching strategies
- Behaviourally focused approaches
- Educational strategies derived from appropriate theory and research
- Provision of adequate time, intensity and materials
- Involvement of parents/family

Depending on the specific circumstances, nutritional/dietary assessment in the school environment provides unique opportunities: These range from direct observation of eating behaviour, the use of surrogate respondents to checking accuracy of reports from menus, plate waste et cetera.[55] On the academic side, mathematics offers a unique, real-life opportunity to practice numerical and other cognitive skills. James and Adams [109] claim that nutrition science and mathematics form a natural partnership, since nutrition incorporates numerous mathematical procedures. In Figure 2.2 mathematics concepts and procedures (such as sorting, classifying, statistics, probability, estimation, rates and proportion) that support integration of nutrition are indicated emphasizing the commonalities between the disciplines.[109]

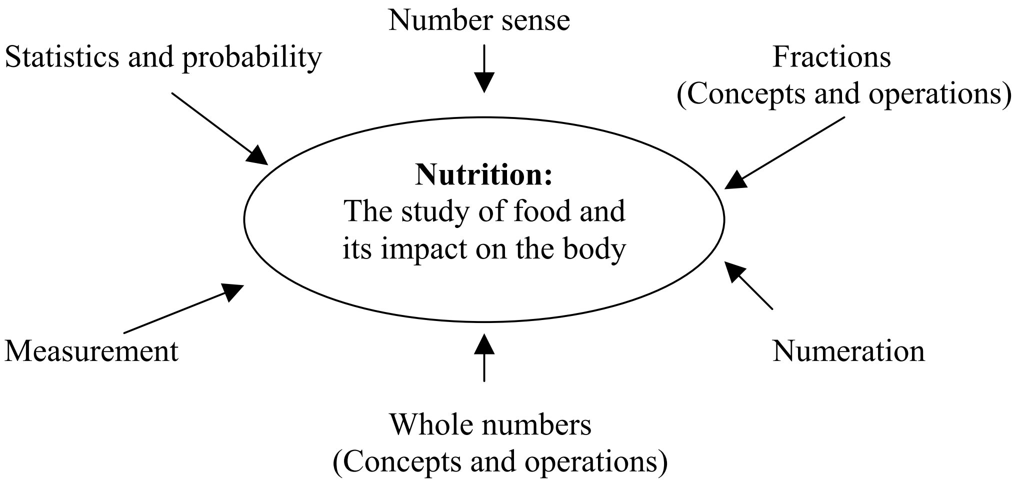


FIGURE 2.2 MATHEMATICS CONCEPTS AND PROCEDURES THAT SUPPORT INTEGRATION OF NUTRITION [109]

According to James and Adams [109] curriculum integration such as linking nutrition concepts to mathematics could have the following specific benefits:

- It encourages learners to use learning experiences to understand themselves and the world in which they live
- It engages the learner in searching for, obtaining and applying knowledge in a non-superficial way
- It provides learners with an opportunity to use their academic strengths to increase achievement
- It allows subject content and relevance to be viewed from various perspectives
- It supports natural, holistic learning
- A broader range of content can be presented in a meaningful way
- It may save time and money for teachers and school administrators

It is thus evident from international literature and from national health and education policy that integrating nutrition into the school environment is not only feasible, but also desirable.

2.2.4 The food frequency questionnaire (FFQ) as basic format of the test method

2.2.4.1 Description

A FFQ or checklist assesses dietary intake by determining how often a person consumes a limited number of foods. [25, 26] The original version was published in 1960 and was outlined as a 'short schedule for qualitative classification of dietary patterns'.[110] Kohlmeier and Bellach [111] have characterised the FFQ as being specifically designed to assess variance in the frequency of intake of particular foods, using a minimal number of closed questions. A respondent is presented with a list of foods or food groups (the item list), and then has to indicate how many times a day, week, month or year (s)he usually consumes these foods.[26] This requires that respondents add up frequencies of consumption across foods and consumption of individual foods across meals. [98, 111] Details about the FFQ are presented in most nutrition assessment textbooks (for example references [26, 27, 110]).

In some FFQ's a choice of portion sizes is not given; a ‘serving’ (that is standard portion size) is established from large-population data and assumed to be true. This is known as the simple or *non-quantitative or qualitative* food frequency questionnaire.[26] Hammond et al [112] used this approach for assessing dietary intake in five to eleven year old children. The *semi-quantitative* FFQ gives respondents an idea of portion size and requests that frequency of intake is provided in terms of this given amount.[26] An example is the widely researched Willet Questionnaire from

Harvard University.[113] Finally, the *quantitative* FFQ asks the respondent to describe the size of his or her usual serving as small, medium or large relative to a given standard. The Block Questionnaire from the National Cancer Institute is an example.[44]

2.2.4.2 Aims

The developmental aims of the quantitative (that is the most detailed) FFQ were: [44]

- Ranking of individuals by relative levels of nutrient intake and also estimation of absolute level of nutrient intake
- Representation of an individual's usual diet
- Relatively brief (for large scale use)
- Capable of assessing nutrients as well as food or food groups
- Assessment of a broad range of nutrients
- Assessment of a variety of demographic groups.

Since the original FFQ's were developed many variations have been designed and the FFQ is now the most popular dietary assessment tool in epidemiology. The comprehensiveness and detail contained in FFQ's vary greatly with regard to the item list, the nature, extent and time frame of the response for recording frequency of intake, and in terms of the reference portion size (from no quantification to sex and age-specific reference portion sizes). Furthermore, some FFQ's assess overall diet, whilst others are nutrient specific [26] and some assess dietary patterns. Consequently most FFQ's now stipulate own aims and usage criteria.

2.2.4.3 Strengths and limitations

The many variations of the FFQ may make generalisations regarding strengths and limitations invalid, but an outline of possibilities is given in Table 2.1.

TABLE 2.1: STRENGTHS AND LIMITATIONS OF FOOD FREQUENCY QUESTIONNAIRES (based on references [25, 26, 110, 114])

Strengths	Limitations
• Can be interviewer or self-administered (relative simplicity) • Trained interviewers not needed • Can be machine readable (if pre-coded) • Modest demand on respondents resulting in increased compliance (compared to other 'traditional' methods of dietary assessment) • Relatively inexpensive for large sample sizes • An indication of usual dietary intake may be obtained • Design can be based on large population data • Considered by some as the method of choice for research on diet-disease relationships (epidemiological studies) • Suitable for ranking or classification according to nutrient intake • Procedure does not alter habitual dietary habits Total diet or selected food or nutrients can be assessed • Response rates usually high (low respondent burden)	• Recall depends on memory • May not represent usual foods or portion sizes chosen by respondents because of incomplete or inappropriate listing of foods and errors in quantification • Intake data may be compromised when multiple foods are grouped within single listings • Not appropriate for determining absolute nutrient intakes like NHANES III • Development and validation is difficult, tedious and may be expensive • Limited data in terms of food descriptions • Period of recall imprecise • Respondent burden is governed by number and complexity of item list and quantification procedure • Recall of past diet may be biased by current diets • Heterogeneity of populations influences reliability of method (suitability questionable for segments of population consuming atypical diets or foods not on list) • FFQ with long list tend to overestimate and those with short lists tend to underestimate intake • No information on meal patterns throughout day • Considerable programming time and expertise required to convert food frequencies to nutrients

2.2.4.4 Development of food frequency questionnaires (FFQ's)

The basic design questions, which need to be addressed when developing a FFQ, include the following: [111, 114, 115, 116]

- Is information needed on foods, nutrients, dietary supplements or other food constituents, or specific behaviours?
- Which foods should be included?
- Which foods should be grouped together?
- Which consumption frequencies should be allowed?

- Is amount of consumption required? If yes, what should be set as a usual portion size
- How should individual foods be weighed in the development of the nutrient database?
- What should be the reference period?
- Is absolute or relative intake needed?
- What level of accuracy is required (individual versus group information for reproducibility and validity)?
- What are the constraints in terms of money, time, staff, and respondent characteristics?

Since cognitive psychologists became involved in FFQ construction, more design questions have evolved, resulting in studies investigating and designing the FFQ's and other dietary assessment methods from that perspective.[89, 117, 118]

Ideally, a FFQ's item list (that is the first three points in the above list) and the associated quantification (that is frequency of intake, portion size estimation and nutrient database) should be based on food intake data representative of the target population [44, 111, 112, 115, 119] in order to address content and measurement validity.

When Block et al [44] developed their FFQ (even though they then called it a self-administered ‘diet history’), they identified two fundamental questions that determine the performance of any FFQ:

"How accurate can the individual report on his frequency of consumption and his portion sizes?"
"How adequate is the food list itself, and its associated quantification?"

The latter is totally the responsibility of the investigator, and consequently instrument development, developmental evaluation and refinement are of critical importance to ensure the potential of precise measurement. Thus, during instrument development a methodological rationale aimed to address a limited, but essential question, namely: *"If respondents were able to respond accurately about their diet, could this food list and associated quantification adequately represent individual dietary intake, in spite of the diversity of dietary behaviour?"* should be provided.[44]

FFQ's should be adjusted for the population group for which they are intended,[116] since the context may well affect the participants' responses and validation characteristics.[120] The key questions to be considered before applying a FFQ designed for another population are the following: [111]

- Does the tool capture 80-90% of the interpersonal variance in consumption for the foods / nutrients under study? To do so, the questionnaire must be based on or compared with a recent survey of total diet assessed independently. This information must be conducted in the age, race, gender, ethnic, or religious group under study.[111]
- The categories (response options) of frequency of consumption also need to be examined for appropriateness for the group understanding.[111]
- Do the nutrient values attributed to each response apply in this group? Individuals consuming pumpkin pie will have incorrect carotene intakes if the nutrient data are based primarily on apple pie consumption. The nutrient value assigned to the questions should be a weighed average of up-to-date nutrient information on all consumed items that the question subsumes, because foods differ in some or all of their dietary constituents.[111, 121]
- Are there systematic biases in response between groups of people of interest? Is the accuracy of information captured, for example, in 50-year old African-American men similar to that from 50-year-old Caucasian men? Do 70-year-old men respond with a different degree of errors to the same questions than 20-year old men do? If so, applying the same FFQ to all people in studies spanning such different groups can result in an artefactual effect of diet.[111]
- Are the assumptions about portion size appropriate for the subjects under study in terms of their gender, age group and population? Are a significant number of participants in the study (for example elderly, vegetarians, children or members of a particular culture) for whom the assumptions on portion size are inappropriate? Use of a single set of portion sizes could then result in over- or underestimation of intakes.[111]
- If the FFQ is used to monitor changes in intake over time, how will change be accounted for? How will the introduction of new foods into the market, changes in price, changes in the use of specific food ingredients (for example oils used in manufacture of margarine) be handled?[111]
- Should a separate dietary assessment tool or biomarker be administered in a sub-sample of the population to calibrate the results from the FFQ and adjust for errors?[111]

2.2.4.4.1 Item list

Constructing the item list is the first step in the development of a FFQ and involves the decision which foods to include and which foods to group together. The aims of the FFQ will be the determining factor when the decision regarding which foods or food groups to include is made, and, as mentioned above, it should ideally be based on a recent survey of dietary habits in the target population.[111]

Since it is impossible to ask about all foods eaten, grouping is a core decision in the development of a FFQ. Block et al [44] used the following criteria for grouping or keeping foods separate:

- Conceptual similarities
- Respondents' ability to make the necessary distinctions
- Similarity in nutrient content per usual serving (not per 100g)
- Importance of a particular food to researchers' ability to correctly classify an individual with respect to nutrient intake
- Approximate number of persons at risk of such misclassification

The reductionism and summation, which are the logical consequence of the grouping, may lead to some loss of the real variance between subjects (in terms of individual foods consumed) and at the same time introducing between-subject variance (in terms of the food groupings).[122]

Little is known about the way in which children deal with individual foods or groupings, but Koehler et al [92] suggested that children more accurately recalled specific items rather than categories in their ‘Yesterday's Food Choices’ instrument.

2.2.4.4.2 Quantification

Quantification of a FFQ includes the portion or serving size attributions of each line item and also the assumed nutrient content of each. In addition, the measurement of frequency of intake should also be considered since all of these eventually determine the quantitative relevance of the result.

- ***Reference portion size and portion size estimation aids (PSEA)***

Terminology: A *portion* is the amount of food that a person reports as being the quantity usually consumed. There is no standard portion size and no single right or wrong portion size. By contrast, the terminology around *servings* originated in the USA and is a standard amount used to help give advice about how much to eat, or to identify how many kilojoules and nutrients are in a food. Serving sizes are specified for the Food Guide Pyramid, Nutrition Facts Label (on food packaging) and the Exchange Lists for Meal Planning.[123]

Criteria for serving sizes in the Food Guide Pyramid are: [123]

- Amount of food from a food group typically reported in surveys as consumed on one eating occasion
- Amount of food that provides a comparable amount of key nutrients from that group (for example amount of cheese that provides the same amount of calcium as 1 cup of milk)
- Amount of food recognised by most consumers (that is household measures) or that can be easily multiplied or divided to describe a quantity of food actually consumed
- Amount traditionally used in previous food guides to describe servings

Food label servings are defined by the US Food and Drug Administration as 'Reference amounts customarily consumed per eating occasion' and are stipulated in the 1990 'Nutrition Labeling and Education Act'. In South Africa the 'Foodstuffs, Cosmetics and Disinfectants Act (1972)' regulates labeling, but manufacturers define the serving size.

Exchange list servings are specific amounts of food that contain about the same amount of carbohydrate, protein and/or fat and energy as other foods on the same list. Serving sizes of different foods on the same list vary.[123]

Even in the USA the public appears to be confused by the terminology. In a survey amongst grade three to five year old children they were reported to use the words 'serving' or 'helping' to name a food portion.[85]

<u>Portion size in FFQ</u>: The quantitative and the semi-quantitative FFQ include an estimate of usual portion size consumed. In the consensus document for the use and development of FFQ [114] it has been stated that allowing subjects to estimate portion size in the completion of FFQ's is considered more advantageous than using average portion size. Suitable methods are the use of defined small / medium / large options or estimation of portion size using photographs.

It has been shown that people have different concepts of medium. Therefore portions need to be defined in quantitative (preferably weight) terms rather than qualitatively.[124] In Western societies a 'portion distortion' is common. Portion sizes both inside and outside the US home are increasing and super-sized portions are becoming the norm in fast food establishments [125, 126, 127] implying that medium portions as used on food labels and by health professionals need to be revisited.

Estimating portion sizes is a challenging task for respondents. Therefore the use of PSEA has been introduced,[128] including two-dimensional (2D) (drawings of real food, abstract shapes and household measures; photographs; graphics; package labels) and three-dimensional (3D) (household measures; real food samples; models; replicas; bean bags and cartons of varying sizes; common objects, for example domino, card deck, palm of hand, fist, tennis and golf balls) aids.[129, 130] A review on the validity of PSEA indicated that, overall, differences between the different types of PSEA were not statistically significant.[129]

Participants when judging portion sizes use many cognitive strategies. Furthermore, for different foods different PSEA may be most appropriate. Qualitative research suggests that respondents preferred aids that were similar in size and shape to actual portions consumed for liquids or amorphous foods, and for solid foods they preferred a ruler.[131] For snacks, bowls seemed to provide a means by which individuals could accurately estimate their consumption.[132] In cases where foods are irregularly shaped, for example muffins, estimation strategies that did not use PSEA were more effective and reference to ‘large’, ‘extra large’ or ‘jumbo’ was a preferable approach.[131] In a study involving an African population it was found that solid foods were better estimated than food with an amorphous appearance when using a food portion photograph book as PSEA.[133]

Training and exposure to PSEA seem to result in improved reported portion size accuracy but a great deal of estimation error remains.[130, 134, 135, 136]

Limited research included children. Goodwin et al [137] found that ten to twelve year olds were able to use portion size model booklets as an adjunct to food recording, but the children reported that they would be reluctant to use it when their peers were around. Frobisher and Maxwell concluded that for subjects 16 years and younger alternatives to the food atlas and descriptions would be more appropriate (for example standard food portions) as they observed high error rates.[138] Assignment of a standard serving size was also recommended for dietary assessment of young children where caregivers acted as data source.[139]

- ***Frequency of intake***

In most cases the main determinant of variation in measuring dietary intakes is frequency of consumption of the food items in the list,[140] but in certain contexts (for example rural areas of Korea) between-persons variation of food such as cooked rice might be determined by portion size rather than by food frequency.[141] The frequency of intake on a FFQ can be detailed, that is requesting that a participant indicates the number of times each line item is consumed in a specific time period (per meal, day, week et cetera) or the response options can be categorised, meaning that a participant has to check intake in limited, preset categories typically in a fixed time frame (for example per week).

The cognitive tasks involved in providing a correct response on a categorised quantitative food frequency questionnaire starts off with recognizing the items on the list.[89] This is followed with at least the following (based on reference [88]):

- Recalling own intake
- Counting number of times per week foods are eaten
- Summation of counts of foods belonging to a food category
- Conversion of summated intake to a frequency of intake format used in the data collection tool (that is coding and recording the answer in the required questionnaire format).

Smith reviewed the cognitive psychological aspects of relevance to reporting FFQ's, including the credibilty of frequency judgments, the difference between rates versus counts and the cognitive implications of item grouping. It was concluded that many of the assumptions underlying the use of the FFQ should be subjected to scrutiny.[89]

- ***Nutrient content of items***

The nutrient database consists of the foods that have been determined to contribute the most to the variance in intake of the nutrient of interest in the population under study.[111]

Nutrient calculations for FFQ's are based on specially constructed nutrient databases that apply weighted averages of the proportions of intakes of all foods covered by the questions. Determination of average nutrient values requires up-to-date information on the relative consumption of the individual items as a proportion of the group under question in the population of interest. For example, if middle-aged African-American men eat more pork than Caucasian men of the same age, the values for thiamin for the response of daily consumption of beef, pork, or lamb will be based on a different database that weighs pork more heavily that the amount for a daily response for Caucasians. These assumptions about relative consumption need to be checked regularly and updated if necessary.[111]

Some of the methods used to assign nutrient values to food groups in FFQ's have been evaluated. In general, mean-based methods appear to be superior to median-based methods, but among the mean-based methods no one variation was consistently better.[121]

2.2.4.5 Food frequency questionnaires (FFQ) for children

When the FFQ is used for school-aged children it should be adapted. The following should be addressed: [142]

- The food list - If composites are used it is important to make sure that the children know which items are included
- Time interval - Time intervals need to be fixed by meaningful start and end points, and may need to be abbreviated
- Response set - Children may respond affirmatively to authoritatively phrased questions or they may adopt a response set when they are unsure, have no opinion, or are disinterested
- Context of questioning - Language needs to be consistent with the child's understanding
- Structuring of the questionnaire - Begin with easy questions on topics of interest with threatening or difficult questions last

McPherson [25] reviewed a total of 21 studies where the FFQ was validated in school-aged children. Of these twelve focused on specific food groups or nutrients and nine assessed the general diet. Parents served as proxies for children in six studies and assisted children in five studies. In one study [48] the accuracy of children's versus parents' responses was measured. The FFQ provided a better appraisal of eleven to twelve-year old children's intake when administered to parents rather than the children: overestimation of energy was more severe for children than for parents. No consistent patterns emerged for either age or gender.

Bellu et al [119] validated a 116 item FFQ completed jointly by Italian parents and their nine to twelve-year-old children by comparison to a seven-day dietary record kept by the parents. The overall validity of individual estimates was “fair for some but unsatisfactory for many nutrients”. The FFQ tended to overestimate intakes but calculating nutrient densities reduced the difference. When the FFQ was compared to 24-hour recalls obtained from the mothers the performance improved on average.[143] This highlights the challenges regarding the choice of a reference method when no 'golden standard' is available.

In a Belgian study the test-retest reproducibility and relative validity of a 15-item FFQ was assessed in three separate studies involving school-aged children aged eleven to 18 years.[144] The FFQ was found to be sufficiently reliable and valid to be useful for ranking subjects and could thus be used for multivariate or correlation analysis in epidemiological studies, but not for estimating prevalences.

In a small (n=20) sample of Swiss adolescents aged nine to 19 years Cavadini et al [145] found that a semi-quantitative FFQ administered at school during a one-hour session "correctly described food consumption" when compared to a modified diet history. Energy and macronutrient intakes estimated from a FFQ (Block98) and a three-day diet record in young girls (four to nine years of age), primarily by their parents, disagreed. It was found that the FFQ overestimated intakes.[146]

2.2.4.6 Food frequency questionnaires (FFQ) for fat intake

Dennis et al examined the quality of FFQ's aimed to assess the relationship between fat intake and prostrate cancer. A total of 39 studies met their inclusion criteria. From these studies they compiled a scoring method for evaluating FFQ's in general. They concluded that, whilst the FFQ has often been used to measure fat intake, methodological flaws in instrument development might partly be responsible for the inconclusive results regarding the relationship between dietary fat intake and prostate cancer.[147]

Studies in which the validity of the FFQ for specifically measuring fat intake was studied include the following:

- A 104-item FFQ aimed to assess fat and cholesterol intakes was found to correlate with biomarkers (linoleic acid in erythrocytes and adipose tissue) and a diet history in 191 adults.[148] Additional evidence that a FFQ can provide informative measurements of dietary fat was published by Willet et al,[149] who used plasma fasting triglyceride levels as an "alloyed gold" standard.
- Validity and reliability of a self-administered FFQ, designed to be sensitive to low-fat, regional and ethnic dietary patterns, were adversely affected when it was administered to minority or poorly educated populations.[150]
- A semi-quantitative FFQ did not provide reliable estimates of actual absolute or percentage fats or cholesterol in subjects consuming diets of known composition. [151]

- Incorporating fat-modified foods into the Block FFQ's item list improved the classification of fat intake when four two-day food records over a one-year period acted as reference method.[152]
- A FFQ specifically adapted to measure fatty acid intake was validated against seven-day weighed records and found to be a reliable estimate of dietary intake of individual fatty acids.[153]

2.2.5 The food record

2.2.5.1 Description

In this method of dietary assessment the participant records, at the time of consumption, the identity and amounts of all foods and beverages consumed for a period of time. Food and beverage intake can be quantified by estimating portion sizes, using household measures (that is *'estimated food records'*), or weighing the food or drinks on scales (that is *'weighed food records'*). Certain items, such as eggs, apples or cans of cool drink may be recorded as units or simply counted.

The food record does not depend on memory because the participant ideally records intakes at the time of consumption. This is in contrast to the FFQ where the task is to recognise food on a list and then remember and calculate usual frequency of intake. Thus, the error structure of the food record differs from the FFQ. The food record can provide detailed food intake information about eating habits (for example when, where and with whom food was eaten). Multiple-day data is more representative of usual intakes and non-consecutive, random days (including weekends) covering different seasons are necessary to arrive at useful estimates of usual intake.[26, 110, 154]

2.2.5.2 Strengths and limitations

The strengths and limitations of food records are summarised in Table 2.2.

TABLE 2.2: STRENGTHS AND LIMITATIONS OF FOOD RECORDS (based on references [25, 26, 110])

Strengths	Limitations
• Do not depend on memory • Defined record time • Intake can be quantified • Theoretically more accurate • Can provide detailed intake data • Can provide data about eating habits • Multiple-day data more representative of usual intake • Reasonably valid up to five days • Training can be group-administered • Procedure (technical instructions) can be automated	• Require high degree of cooperation • Recorder must be literate • Response burden can result in low response rates • Take more time to obtain data • Act of recording may alter diet • Data collection and analysis are labour intensive and expensive • Food eaten away from home less accurately recorded

A food record requires that that children can write names of foods legibly, recognise and describe quantities in either fractions or whole units, decipher food label information, and retain the record in their possession for completion of all entries during a day.[155] Children ages ten to twelve years are reliable respondents; adolescents are capable but often less interested in participating than younger children; Standardised procedures for completing the record reduce respondent error and food illustrations or flow sheets assist the recording technique.[155]

Wold et al were able to show that instructional flowcharts helped participants keep three-day records in terms of accurately describing food intake. Such records tended to be complete and specific.[156]

2.2.5.3 The food record as reference method

Some authors have referred to the food record as the 'practical golden standard' in dietary assessment.[110, 157] From the above-mentioned advantages it is evident why, over the years, it has extensively been used as the reference method in comparative validation studies where the even more costly methods such as duplicate portions,[125] direct observations,[158] doubly labeled water,[159] or other biomarkers [160, 161] were impractical or too expensive.

A key issue in validation studies is the independence of errors between methods.[154, 160, 161] From the summary of advantages and limitations of the FFQ and the food record, it is evident that the error structure of the food record is reasonably different from the error structure to be expected from the FFQ, making it a suitable reference method.

In the case of individuals, the food record measures current or actual intake.[162] Thus, if usual (that is habitual) intake of individuals is to be measured, multiple, non-consecutive days covering the reference period (that is all natural and trade / cultural seasons) are recommended, [26, 163, 164] since average intake reported in a long series of food records has been considered an operational definition of 'usual diet'.[157]

An unresolved question regards the number of days that should be recorded in order to assess usual intake.[165, 166] Carroll et al [157] investigated whether it was better to obtain many food records from a moderate number of subjects or a small number of food records from a larger number of subjects when the food record is used as 'gold standard' in a validation study. They concluded that neither strategy is always preferable: The aim of the validation study is of prime importance. For estimating correlations or slopes (as in the present study) within-person variance of the food record and distribution of true usual intake seem to be deciding factors.

The number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence ($P<0.05$) have been published and those of relevance to this study are indicated in Table 2.3. The group size on which Basiotis et al [167] based their findings was relatively small (n=29) and they conclude that fewer days would be needed for larger groups. Overall, energy intake seems to require the least number of days to classify 80% of the population into tertiles of nutritional intake with a 95% confidence interval, [166] and fat seems to have relatively low intra-person, short-term variability and tends to remain stable in healthy free-living people.[168]

TABLE 2.3: AVERAGE NUMBER OF DAYS REQUIRED TO ESTIMATE TRUE INTAKE OF INDIVIDUALS AND GROUPS OF INDIVIDUALS (selected information from reference [167])

Component	Individuals		Groups of individuals	
	Males	Females	Males	Females
Food energy	27	35	3	3
Fat	57	71	6	6
Saturated fat	71	87	8	7
Cholesterol	139	300	13	15

Beaton [122] was able to show that for many nutrients, increasing the number of days of data beyond three does not materially contract the distribution, but it remains important to remember that this recommendation refers to a large sample of subjects and a true distribution of usual intakes. Relatively fewer days of recording appear to be necessary to describe the intakes of younger persons and people who are characterised by diets with limited diversity.[26, 166] However, if the aim of a study is to capture food variety in school children, then a two-week period appeared to be necessary for US fourth and fifth graders.[169]

Another issue in food recording is the decision which days to include. In order to represent all days of the week proportionally, the classical approach was the seven-day-food record.[110, 170] Since longer recording periods may reduce accuracy and cooperation, [171] but weekend days may be different from weekdays, ensuring that weekdays and weekend days are included is typically recommended.[26]

It appears that the variation in diet on different days of the week could be population-specific. Jula and co-workers studied the influence of days of the week on reported food, macronutrient and alcohol intakes in Finnish adults. They found relatively little variation on weekdays (excluding Fridays) but clear differences on Fridays, Saturdays and Sundays. They thus recommended five days of recording including any two days from Monday to Thursday, plus Friday, Saturday and Sunday.[172]

A South African study from the mid 1960's compared Indian, black and coloured children in Pretoria. A rank order of correlation coefficients between daily intakes of twelve nutrients on each day of the week and the average weekly intake revealed the following: Weekend days did not necessarily give a poor representation of a week's average intake, but certain days of the week may give a better indication of average daily intake for the week than others. For Indian children the highest correlation coefficient was found for a Thursday. Weighing food intakes on a Friday gave a more representative picture of the average daily intake of black children for the week than weighing on any other day, whilst for coloured children there was no difference from day to day.[173]

Research has indicated that dietary reporting decreases during the recording period. Therefore it has been recommended that in multi-day records the starting days should be evenly distributed across the days of the week to counter recording fatigue, boredom and training effects.

Otherwise the introduction of a systematic error with repeated measurements could outweigh any advantages of repeated records.[81]

From the above it is evident that many factors influence the decision about how many and which days to record, including the purpose of the study, the sex and age and typical diet of the group to be studied, and the nutrients of interest.[26] In general, it is recommended to include measures of quality control when collecting dietary data.[81]

2.2.5.4 The food record for children

A review on dietary assessment methods among school-aged children included six studies where the validity of the food record was studied.[25] The ages of the children varied from eight years (where adult assistance was required) to 19 years. In general, the authors reported that food records underestimated energy intake when compared to doubly labeled water. Few studies evaluated children's ability to complete the record on their own or to record an entire day.

The following additional studies investigated the validity of the food record in children:

Jenner et al [48] compared various methods of dietary assessment amongst Australian school children and found that two or three-day, carefully administered and thoroughly checked, food records were reasonably valid means of assessing usual intakes in eleven and twelve year olds. They reported consistently higher correlations and smaller differences between food records and their reference method (14 food records collected over several months), compared to any other (food frequency type) reference method tested. Their test and reference method were, however, both food records and 'auto-correlation' between methods that have the same error structure may have played a role, particularly because the two- or three-day record series was included in the 14 day series.

In the NHLBI Growth and Health Study nine- and ten-year old girls were assigned to one of three dietary assessment methods (24-hour recall, three-day record and five-day FFQ). At the same time unobtrusive observers recorded types and amounts of food eaten during lunch. It was found that the nature of errors in food reporting and quantification varied with the assessment method. The three-day record was reported to have a comparative advantage over the others in this age group.[174]

2.3 NUTRITIONAL AND DIETARY SCREENING

2.3.1 Definition and characteristics

In general, the WHO views screening as the use of presumptive methods aimed at detecting unrecognised health risk or asymptomatic disease in order to permit timely intervention (Braveman & Tarimo, 1994:6). This definition is very similar to that quoted by Rush: [175] It is "*the examination of asymptomatic people to classify them as likely, or unlikely, to have the disease that is the object of screening. People who appear likely to have the disease are investigated further to arrive at a final diagnosis*".

The ADA defined nutrition screening as "the process of identifying characteristics known to be associated with nutrition problems", [176] and the Joint Commission for the Accreditation of Health Care Organizations of the US and ESPEN of Europe have formally acknowledged its importance in patient-oriented, high standard nutrition care.[26, 36]

Screening is often the first step of nutritional assessment (level 1 nutrition care) provided by qualified health care professionals in hospitals, clinics, private practice and community settings with the aim to (i) identify children at risk, (ii) refer children at risk to a registered dietitian for in-depth nutritional assessment (level 2 care), and (iii) provide anticipatory dietary guidance and educational materials to families regarding prevention of nutritional problems.[177]

Splett and colleagues [178] consider screening as a trigger event, which initiates the nutrition care process. Thus it forms the access point for referral to nutrition care. The purpose of nutrition screening is to predict the probability of a better or worse outcome due to nutrition factors.[36] It can happen either during a general health or disease-focused screening, recognizing that a nutrition-related need can be identified as a potential or early risk factor or as a complicating or underlying factor related to an existing medical condition or disease. In disease-focused screening, nutrition problems can be a cause of a result of the disease, and they can have a physiologic or behavioural aetiology.

According to the ADA the nutrition screening process is characterised by the following: [176]

- it may be performed in any setting (given the opportunities and constraints of the intended venue, for example hospital-based versus tools for the home dwelling) [175]
- it aids in achieving early intervention goals

- it includes the collection of applicable data on risk factors and the interpretation of data for treatment or intervention
- it determines the need for a complete nutritional evaluation
- is cost-effective.

Rush [175] and Kondrup et al [36] specify that the usefulness of screening tools and programs should be evaluated by assessing thier predictive and content validity, reliability, practicality and link to action protocols. Furthermore there should be significantly greater benefit from earlier intervention than from what would result from intervention at the time the subject seeks help because of symptoms. Finally, screening should be shown to be preferable to other strategies, such as universal application of an intervention.

Nutrition screening can include anthropometric, biochemical, clinical and / or dietary data [26] and, in the case of young children, behavioural and skill development have been added.[177] Since there is no fixed boundary between screening and diagnostic testing [175] any nutritional status indicator can potentially be used for screening. The limitations are set by the complexity, cost and utility of the screening tool or protocol, by the prevalence of the problem being assessed and the potential benefit from the intervention. Thus dietary screening is a form of brief nutritional assessment focussing on food intake.

Interest in dietary screening is based on the observation that a large fraction of the variability of nutrient intake can be explained by a small number of foods.[179] Simple tools designed for dietary assessment (for example HEA1, HEA2, HEA3, DINE, Nurse Questionnaire) were shown to perform as well as much more complicated and time-consuming tools (for example 24-hour recalls; checklists) and their performance may even be comparable to the seven-day record. [120] There have been reports where a FFQ using seven broad categories correlated more highly with reference values than a FFQ using 31 individual fruit and vegetable items.[180] For children, methods of dietary assessment that are perceived as being less burdensome and time-consuming may improve compliance.[81]

Some short assessments are part of larger surveillance programmes, for example the ‘Youth Risk Behavior Surveillance System’ for American adolescents, which includes seven questions about the previous day's food choices.[181]

2.3.2 Aims

In general, there are different types of screening, each with specific aims: [45]

- *Mass screening* involves the screening of a whole population
- *Multiple* or *multiphase screening* involves the use of a variety of screening tests on the same occasion
- *Targeted screening* of groups with specific exposures
- *Case finding* or *opportunistic screening* is restricted to patients who consult a health practitioner for some other purpose

Wilkin et al [182] divide the purpose and use of health-related measures into three broad categories, namely discrimination, prediction and evaluation. According to these researchers there may be more than one purpose for a tool. In the case of *discrimination*, the purpose of the measure is to classify individuals or groups based on some health-related dimension, for example as a means for identifying areas of need, or to help target those whose needs are greatest. The purpose of *prediction* is to identify groups or individuals who have or will develop some target condition or outcome. It is thus aimed at predicting future need at an early stage in order to save time or costs, or to be predictive of a more detailed assessment. *Evaluation* is intended to measure or monitor the magnitude of longitudinal change in individuals or groups on the dimension of interest, for example by focusing on changes over time attributable to an intervention.

Four fundamental measurement axioms have been proposed for any tool used for health measurement. De Vos [183] summarised these as follows:

- If an instrument must have any utility in practice it must be valid and reliable
- For maximum utility an instrument must be brief, easy to administer, easy to understand, score and interpret
- There are only two ways to determine whether a client / patient has a problem: watch him or ask him. Thus direct observation and client report are the methods in which information can be obtained
- There are only four ways of measuring client / patient problem: in terms of its switch, frequency, magnitude or duration. Switch refers to presence albeit absence of the problem. The frequency is obtained by assessing how often the problem is encountered. Magnitude or intensity characterises the degree to which the problem is present, and duration specifies the length of time the problem is continually present.

Each of the above axioms is seen to have general and specific implications for the development of a dietary screening tool as a health measure.

According to Keller et al [184] and Jones [185] the criteria for the tool must be stipulated by specifying what the tool is intended to achieve in which population. In this context it is important to clarify whether the measure focuses on individuals or groups.[182]

2.3.3 Examples of screeners

Some screening tools are intended to pick up general nutritional risk of adults (for example references [186, 187]). A computerised diet questionnaire for the use in health education aimed at giving rapid feedback to the general public was validated against 16 days of weighed diet records. At least 65% of subjects were classified to within one quintile of the classification of the record for most of the nutrients assessed.[188]

Another group of screeners may be aimed at specific target groups and contexts. Examples are the numerous tools that have been developed to identify nutritional risk of the hospitalised patient on admission for example the 'Derby Nutritional Score' [189], the 'Veterans Affairs Nutrition Status Classification' [190], nutritional scores such as the 'Prognostic Nutritional Index' (PNI), the 'Nutritional Risk Index', the 'Subjective Global Assessment' (SGA), the 'Mini Nutritional Assessment' (MNA), the 'Registered Nurses Nutrition Risk Classification' and others.[191, 192, 193]

Several dietery screeners that measure fruit and vegetable consumption have been published.[194, 195, 196, 197]

The 'Family Eating and Activity Habits Questionnaire' is an instrument that identifies factors in a child's family environment that facilitate obesity.[198]

2.3.3.1 Fat screeners

In two publications [199, 200] an overview of short or qualitative questionnaires assessing fat intake has been provided.

A number of fat screeners used or included elements of a FFQ. These include:

The original MEDFICTS instrument as recommended by the NCEP is a simple approach for rapidly assessing a person's adherence to the Step 1 and 2 diets,[26] making it an efficient tool in cardiovascular screening, clinical practice, or research aimed at detecting individuals consuming diets above or below the cut points set out in the 'Heart-Healthy' (Step 1) and 'Therapeutic Lifestyle Change' (TLC) (Step 2) diets. It was originally developed for Adult Treatment Progam (ATP) II and has again been included in ATP III.

In a pilot test validation the MEDFICTS score was significantly correlated with percentage energy from fat (r=0.8, P=0.0002), percentage energy from saturated fatty acids (r=0.8, P=0.0003), and dietary cholesterol (r=0.5, p<0.05).[201] Similar results were obtained when MEDFICTS was self-administered or nutritionist-administered, and then compared to recent three-day food records.[202] More recently Kris-Etherton et al [203] confirmed the validity in follow-up studies. The MEDFICTS dietary assessment tool has also been adjusted to accommodate cultural differences.[204] Taylor et al [205] validated the tool amongst adult army recruits.

Block et al [206] developed a 13-item FFQ type dietary screener for high fat intake. They found that among 101 females aged 45years and older, that the tool performed nearly as well as a four-day diet record in correctly identifying those above and below the group midpoint in PFE. Caan et al [49] modified this tool and evaluated its sensitivity, specificity and predictive value. They found variations in these indicators of validity leading them to conclude that it could not be used as a single assessment method.

A qualitative fat index was validated with a three-day food record as reference method. The index was based on four questions, which reflect the most important sources of fat in the Finnish diet. This was supplemented with a short FFQ consisting of 21 items. The latter proved to be accurate at group level and the former for measuring quality of fat.[207]

The 'Fat List', a short FFQ, was compared to the seven-day food record. For Dutch adolescents the correlation between the two methods for total and saturated fat intake in grams was 0.6. For percentages energy from fat the correlations were low.[199]

Murphy and colleagues developed a food behaviour checklist for use in low-income (EFNEP) groups and evaluated its criterion validity using biomarker and convergent validity with multiple 24-hour recalls. Overall, the fat and cholesterol-related items performed poorly, their internal consistency was low and correlations with PFE were weak.[208]

A behavioural approach or combinations of methods formed the basis of a number of other dietary fat screeners:

The 'Food Habits Questionnaire' of Kristal et al [209] is a behavioural approach to assessment. The 18-item scale had high reproducubility and internal consistency and correlated well with PFE in middle-aged females. Birkett and Boulet [210], however, found poor performance amongst male labourers in terms of reliability (Cronbach's alpha and item total correlations) as well as validity measured with partial correlations.

In the Family Heart Study, a coronary heart disease prevention project, an "inexpensive, reliable and valid" instrument for rapid assessment of eating habits and diet composition was used, consisting of 32 items. The researchers measured validity by comparison with 24-hour dietary recall and by comparing changes in diet with changes in plasma cholesterol levels in a five-year period.[211]

'Rate your Plate' is a brief eating pattern assessment and educational tool used for cholesterol screening and education programmes. The authors stipulate that it is neither a measurement of usual, long-term, nor of quantitative intakes.[212]

The 'Dietary Risk Assessment' (DRA), originally developed by Ammerman et al [213] to identify dietary behaviours associated with cardiovascular disease, was compared with multiple 24-hour recalls and a seven-day recall. The correlations were moderate, but it was recommended as a primary care screening instrument for higher fat intakes.[214] Dietary behaviours related to total fat and saturated fat intake have been identified by Capp et al.[215] The results were expected to have implications for designing brief fat assessment instruments.

A saturated fat / cholesterol avoidance scale consisting of six component items was developed and its internal consistency and criterion validity (relative to scores on the Keys equation and self-report of diet by means of 24-hour recall, FFQ and fat behaviour) were determined. It was recommended as a useful tool in epidemiological research on cardiovascular risk factors. [216]

A twelve-item questionnnaire ('Fat Habits Score') has been developed to evaluate group changes in fat intake. The score was compared in children and adults with estimates of saturated and total fat intake (percentage of total energy) from a FFQ. Both questionnaires were re-administered six months later and it was found that the simple score was able to detect changes in fat consumption (Kinley, 1991).

The ‘Fat Intake Scale’ (FIS) consists of twelve items related to dietary fat, saturated fat and cholesterol. It was compared to food records and the score was found to have acceptable reliablity and validity. The Keys score and the RISCC (ratio of ingested saturated fat and cholesterol to calories) score were additional diet scores with which the FIS correlated.[217]

The ‘Food Behaviour List’ has also been developed by Kristal et al.[218] It is a simplification of the 24-hour recall that consists of 19 yes/no questions about foods consumed the previous day. Its agreement with a professionally administered 24-hour recall was tested. Preliminary evidence suggested that it was a valid measure of lower fat-higher fat intake.

The ‘Diet Quality Index Revised’ [219] reflects adherence to current dietary guidance by populations. Three of the ten components of the tool relate to fat intake (that is total fat =<30% of energy; saturated fat =<10% of energy; dietary cholesterol <300mg).

From the above it is evident that numerous screeners for dietary fat intake have been published. Very little attention has been paid to dietary fat screeners for children.

2.4 RELIABILITY AND VALIDITY IN DIETARY ASSESSMENT AND SCREENING

One of the requirements of dietary assessment and screening tools is that they should be reliable and valid. In the quantification of reliability and validity a distinction should be made between variability and error.

2.4.1 Variability and error in dietary assessment

2.4.1.1 True variability

Dietary intake is characterised by a ‘true variability’, which includes both intra- (within) and inter- (between) subject variation. Since this variation characterises true usual intake, no attempt should be made during the measurement of diet to minimise this variability.[220] Instead, researchers are encouraged to design their projects in such a manner that these two sources of variability can be separated and estimated statistically. In this way the magnitude of the effect of intra- and inter-subject variation can be taken into account during the interpretation of the data.[220]

Variability within the individual may (i) occur from day to day (that is diurnal variation, for example the day of the week effect in dietary intake),[220] (ii) follow a consumption curve (for example natural and commercial seasonality)[163, 221] or (iii) progress with normal growth and development. Gibson[220] adds to this a training effect where a subject alters intakes in reaction against repeated interviews.

This intra-subject variation is particularly important if data on usual intakes are to be correlated with other parameters (for example biochemical or clinical findings), since large intra-individual variation in intake will tend to reduce the absolute value of the correlation. The resulting attenuation of the correlation coefficient could, for instance, be a reason for the apparent lack of a significant relationship between dietary fat intake and serum cholesterol levels in individuals.[220]

Variations within populations (inter-individual variability) can be considered the cumulative variability of individuals, and, generally speaking, knowledge about variability in populations makes it possible to define ranges of 'normality'. Environmental (for example geographic) and genetic influences play a part in this regard. In addition, age and gender are sources of inter-individual variation that need to be considered in measuring diet.[220]

If inter-subject variation is large relative to intra-subject variation, subjects can be readily distinguished so that usual nutrient intakes of individuals can be characterised. However, for most nutrients, inter-subject variation is smaller than intra-subject variation, and consequently, mean intakes of groups can be measured more precisely than individual consumption.[220] The ratio of intra- to inter-individual variance is nutrient-specific (for example when based on 24 days of records, the within- to between person variance ratios ranged from 1.4 for saturated fats to 4.6 for vitamin A).[111] Similarly, the precision estimates from one 24-hour recall in estimating energy intake for a typical male would be ±51%, whereas it is ±293% for vitamin A.[26] Gender, age, ethnic group, and country are also known to affect the ratio of intra- to inter-individual variance.[26, 111]

It follows that knowledge of the true variability of the attribute of interest (in this case habitual fat intake) is important in order to ensure measurement of the true picture of usual consumption in a particular individual or population.

2.4.1.2 Error

Whilst true variability should be reflected by dietary assessment, measurement errors (which can be due to poor calibration of the instrument, inherent lack of precision of the instrument, or mistakes in the collection, reporting and recording of information by the subject or researcher/dietitian) [222] should be controlled and minimised.[220] Errors associated with the compilation of nutrient data and the nutrient analysis of food items are another source of error, which should be kept in mind.[27]

Two types of measurement error can be distinguished in the measurement of diet: systematic and/or random. [91, 166, 220, 222]

Systematic error occurs when there is a tendency to produce results that differ in a systematic manner from the true values, that is a systematic under- or overestimation in an individual or groups of individuals.[122] It is formally defined as "any process at any stage of inference, which tends to produce results or conclusions that differ systematically from the truth." [223] A study with a small systematic error has a high accuracy, independent of by sample size.[45] Since systematic errors reflect bias, the control (and ideally elimination) thereof should be addressed during the testing and validation of a technique, because they cannot be removed by subsequent statistical analysis.[220]

Over 30 types of specific types of systematic errors have been identified in epidemiology. The two most important examples are, according to Beaglehole et al [45], selection and measurement (classification) bias. Selection bias occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not. Beaglehole et al indicate that measurement bias occurs when the individual measurements or classifications of disease are inaccurate, for example when different laboratories produce different results on the same specimen. Confounding (which arises when the non-random distribution of risk-factors in the source population is also present in the study population) is sometimes added to the systematic errors even though it is not the result from a systematic error in research design.[45]

In nutrition epidemiology selected examples of bias that may apply to assessment of exposure are summarised in Table 2.4. [223] The examples in the Table are errors due to the respondent or the interviewer.[27, 91]

TABLE 2.4: BIAS THAT MAY OCCUR IN NUTRITIONAL ASSESSMENT (based on reference [223])

Type	Description	Comments or examples
Insensitive-measure bias	When outcome measures are incapable of detecting clinically significant associations	May reflect difficulties in accurate recall, portion estimation, and generalisation to 'usual diet'; Relevant to recall methods and particularly FFQ; Unintentional [88]
Underlying-cause bias (recall bias)	Cases may ruminate about possible causes for their illness and thus exhibit different recall to previous exposure than controls	In case-control studies where diet is assessed retrospectively
Unacceptability bias	Measurements which embarrass or invade privacy may be systematically refused or evaded	Obese subjects may be prone to this type of bias
Obsequiousness bias	Subjects may systematically alter responses in the direction they perceive desired by the investigator	In face-to face interview situations; The risk of intentional wrong answers increases if the subject believes that a quality scale is involved, for example the 'desirable' responses of FFQ maybe perceived to be either on the left or the right side of the form [88]
Expectation bias	Observers may systematically err in measuring and recording observations so that they concur with previous expectations	In interviews where unusual diet is reported; Following an intervention, participants bias their reports to appear in compliance with the intervention goals [224]
Exposure-suspicion bias	A knowledge of the subject's disease may influence both the intensity and outcome of search for exposure to the putative cause	When interviewer is not blinded
Attention bias	Subjects may systematically alter their behaviour when they know they are being observed	During food recording diets may (intentionally) be simplified, 'unhealthy' foods avoided or 'healthy' choices increased [225]

Berg et al [226] investigated selection and response bias is a dietary survey of Swedish children in fifth, seventh and ninth grades. They found significant differences between participants and non-participants with respect to socio-demographic and food variables, despite great efforts to obtain a high response rate. A decline in recorded foods during the recording period was also observed. They conclude that these two types of bias are likely to be present in dietary surveys involving children, and consequently this should be taken into account during the planning, analysis and interpretation of data.

Buzzard and Sievert [32] list identifying and minimising bias, particularly non-response bias and other sources of error as research priorities in dietary assessment methodology.

Randomisation, restriction, matching, stratification and statistical modeling are methods to control confounding.[28, 45]

Random error is the divergence, due to chance alone, of an observation on a sample from the true population value, leading to lack of precision in the measurement of an association. It cannot be entirely excluded, yet quality control procedures during each stage of the dietary assessment can increase the reliability and hence the precision.[45, 220] In general, individual biological variation, sampling error, and measurement error are the major sources of random error.[45] Apart from the above-mentioned quality control measures, adequate sample size or taking the average of multiple reference measurements (dietary recalls or records) per subject are the best ways to reduce random error in dietary surveys. This has recently been reviewed by Volatier et al.[166] Formulae for calculating sample size and repeat measurements are available, but cost considerations always play a role.[166]

Systematic and random errors may each occur at the intra- and inter-individual level, [222] the characteristics of which may be summarised as follows (based primarily on text provided by reference [222]):

TABLE 2.5: MEASUREMENT ERROR IN DIETARY DATA

	Random	Systematic
Intra-individual	• Reflects day-to-day variation above and below the individual's true long-term intake • Is the major source of error in dietary data • The magnitude varies by nutrient: Macronutrients vary less, because they make a large contribution to total energy intake; Micronutrients vary more because they are often concentrated in certain foods, and their intake is strongly influenced by food choices for the day • The effect of this error is to attenuate the strength of association, causing the correlation or regression coefficients to be biased toward zero • Methods are available to adjust for this error, provided replicate measures of diet are available. Examples include reliability ratio, correction factors, within-person variance, all of which can be described as "approximation of results that would otherwise be obtained if the estimates of long-term diet were available"	• Best typified by under and over-reporting of intake by some individuals (for example underreporting of energy and fat intake by overweight subjects) • Depends on the accuracy of the reported intakes by the subject or the interviewer and the detection of misreporting and especially underreporting [166]
Inter-individual	• Caused by using only a few measurements per subject in the presence of random within-person error	• Results from systematic within-subjects error that affects subjects non-randomly, for example using incorrect nutrient composition values for some foods may appear to affect all individuals in the same direction, but their impacts are not the same, since consumption of these foods is likely to differ among subjects • Biases in national representative dietary surveys can be linked to non-responders, since the non-responders may differ significantly from those participating in a study. Affected by control of missing or undefined data, the description of foods, procedures to code and aggregate single food items, and data check procedures [166]

Some researchers (for example reference [111]) avoid the use of the terms 'systematic' and 'random', since they reason that systematic errors can be randomly distributed. They propose a distinction between unbiased and biased methods, and differential and non-differential errors.

Based on this reasoning a measurement X' is defined as an unbiased measurement of X if the average measurement approaches the true measure as the sample size increases. Unbiased measurements result for $X' = X + \varepsilon$, where ε is a random error variable with expectation 0. On the other hand, X' is a *biased measurement* when the average measurement does not approximate the true intake, and ε from the equation does not equal 0.[111]

Both biased and unbiased errors can be either differential or non-differential. Measurement error is *non-differential* as long as the error distribution is identical for all individuals of a study or for each subgroup of a population. In the above equation this would occur if the distribution of the error variable ε would be the same for every individual under study. Measurement errors are *differential* if the participants of a study react differently to a measurement method that is used within a study.[111] For example, if hospital patients report a lower variance in their diets than controls, this may be reflective of their current diets, but not the true variance of their usual diet - unbiased, but differential error - because the variance and therefore the errors are differential. If, however, obese individuals underreport their fat intakes, whereas lean subjects report accurately, then this would constitute biased differential error.[111, 122]

The distinction between differential and non-differential measurement error is important for error assessment, adjustment, and correction strategies.[111] For non-differential errors the direction of influence on an estimated exposure-disease relationship is presumed to be biased toward zero. Differential errors influence exposure-disease relations in ways that can only be predicted if information about error in all subgroups under study is available.[111, 122]

It is not possible to statistically distinguish random errors from true intra-subject variation as described in the introduction,[220] unless replicate measures of diet are collected and these replicates are time independent, meaning that the replicates are taken on at random, preferably non-consecutive days.[222] Alternatively, observed intra-subject variation represents the sum of true variation plus the remaining sources of random measurement errors.[220]

The goal of a study or the aim of a dietary assessment tool will dictate the required accuracy. Equally, the impact of measurement errors on the design of a dietary assessment method depends on the aims to be achieved by the study. For example, the larger the random error, the greater the size of the sample required for estimating mean nutrient intake of a group. Also, increased random error increases the number of replicate measurement days necessary to define the distribution of usual nutrient intake of an individual.[111, 220]

The sources of error related to data collection and recording are of prime interest in this study. Intake is recorded as foods. On this level of error (also called ‘misrepresentation’, ‘false memory’, ‘misreporting’ or ‘distortion’), [91, 94, 95] the following can be distinguished:

- *Phantom foods* are items reported but not observed eaten. These are also called ‘intrusions’, ‘commissions’ or ‘false positives’.
- *Omissions* are items not reported eaten but observed eaten, also called ‘false negatives’.
- *Elaboration* is the process of unintentionally distorting information, for example when the participant uses general (‘generic’) knowledge of his / her diet to substitute for the true past information. The phenomenon is also called ‘substitution’.
- *Matches* are foods reported eaten and observed.

If the participant accurately remembers eating a food item, or erroneously assumes that it was eaten, then the quantity consumed must be estimated. This information may not be salient for most respondents. Estimating usual amount is a complex cognitive task, since:

- Food frequency and portion sizes are not constant over time.
- Inferences and calculations must be made when the respondent’s frequency and portion size differ from those given.
- Respondents often do not pay attention to frequencies and portion size when eating.
- Respondents do not have clear mental images of portion sizes eaten when estimating consumption.[88]

Error can also be associated with the compilation of nutrient data. The following is of importance in this regard: [27]

- Inadequate sampling procedures
- Inappropriate analytical methods
- Errors in analytical methods
- Lack of standardised conversion factors for calculating energy and protein content
- Inconsistent terminology
- Incorrect description of individual food items
- Inconsistencies from genetic, environmental, food preparation and processing factors

Kris-Etherton & Etherton [227] and Stumbo [228] emphasised the importance of a current nutrient database, since new foods enter the marketplace and existing foods undergo compositional changes, particularly in the case of fatty acids. In South Africa it was found that fatty acid intake from seven-day dietary records differed when the 1986 Medical Research Council Food Composition Tables were used for analysis compared to the 2000 SAFOODS.[229]

From the above it is apparent, that variability and error are inherent to dietary assessment. Any attempt to evaluate dietary intake needs to take this into consideration. Since underreporting of energy intake appears to be the most common form of misreporting, an analytic framework of predictors of accuracy of reporting has been proposed by Tooze et al (Figure 2.3).[230]

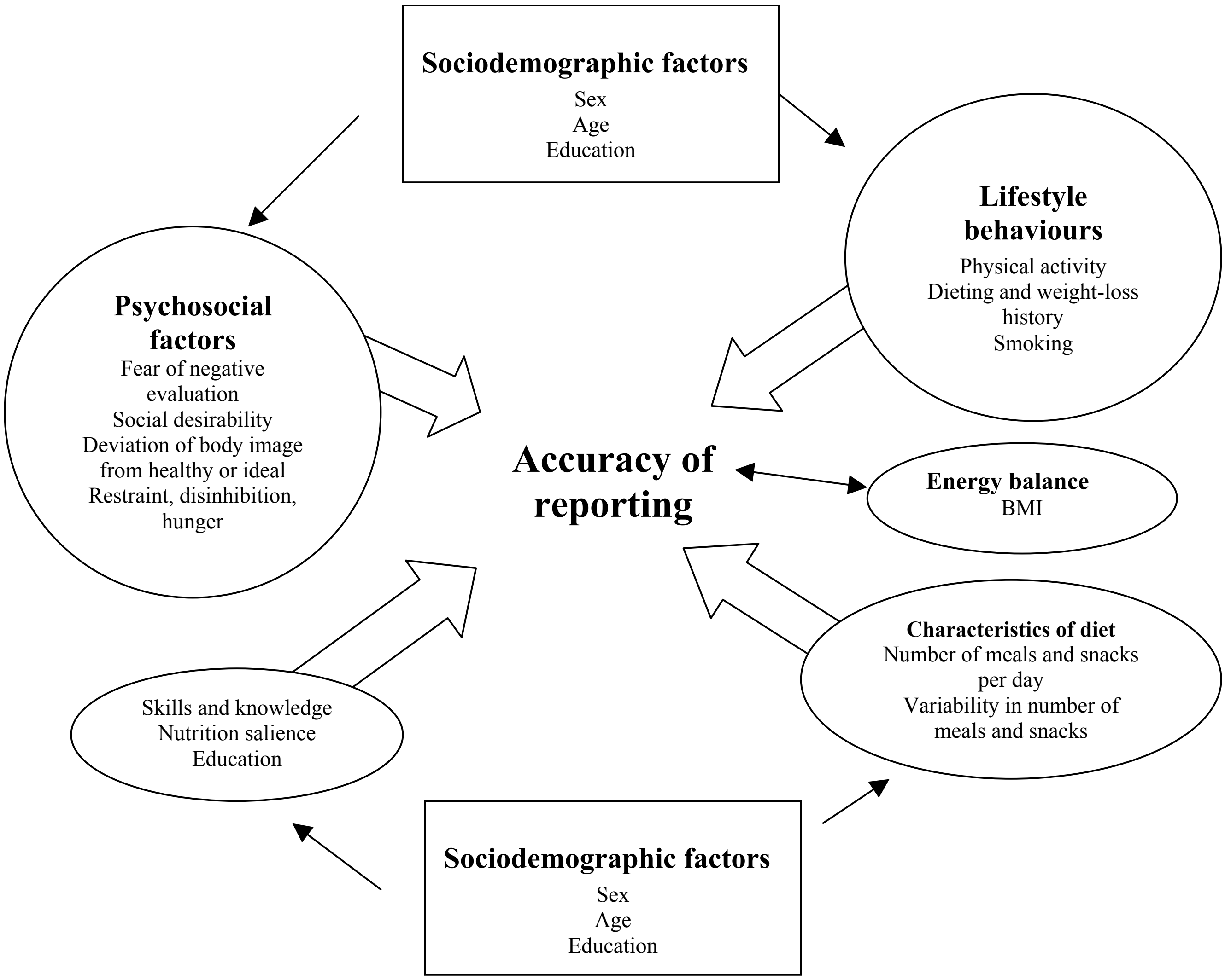


FIGURE 2.3: ANALYTIC FRAMEWORK OF UNDERREPORTING OF ENERGY INTAKE (from reference [213])

2.4.2 Reliabilty

2.4.2.1 Definition

Reliability of measurement (or the lack of it) is the extent of unsystematic variation in the quantitative description of some characteristic of an individual when the individual is measured a number of times.[231] It is a central consideration of validity concerning the process of data collection.

Marais and Mouton [232] as well as Miller and Achterberg [233] state that reliability refers to the requirement that the application of a valid measuring instrument to different groups under different sets of circumstances, should lead to the same observations. Thus reliability could be defined by posing the following question: "Will the same methods used by different researchers and/or at different times produce the same results?" even though Joachim [221] has argued that the term does not have a universal meaning to all researchers, and that the definition "stability of an instrument and measurement process when it is applied under standard conditions" may be problematic in dietary studies, as the meaning and relevance of 'standard conditions' are not clear. They prefer to define reliability as "the ability of an instrument to consistently measure what it aims to measure".

Wilkin et al [182] state that the more reliable a measure is, the lower the element of random error. Unreliable measures cause problems, when the aim of the tool includes the following: [231]

- Comparison or ranking among individuals by means of the measure
- Assigning individuals to groups based on scores obtained in the measure
- Prediction of other traits
- Assessing the (systematic) effects of other factors on the measure

2.4.2.2 Types of reliability

A number of types of reliability are, according to Wilkin et al,[182] of importance in the assessment of instruments:

(i) Consistency over time (Test-retest reliability)

This type of reliability is also called 'reproducibility' [80] or test-retest 'stability',[231, 233] and refers to the ability of an instrument to produce the same estimate on two occasions, assuming nothing has changed in the interim.

Thus, the measure is administered twice to the same group in a relatively short time, under the same conditions.[182, 184] This will determine

- whether respondents interpret questions in the same way during the first and second administration and
- whether the tool is 'stable' (that is answers remain the same despite mood changes or other recent, unrelated events).[184]

Wilkin et al [182] have identified two problems with this approach:

- There may have been real changes in the population between the two administrations. The magnitude of this problem depends on the length of time which lapses between administrations, and the stability of the variable being measured. Consequently, although the results from two separate dietary assessments disagree, the method may not be imprecise: the food intakes may indeed have changed. [27]
- Subjects may either undergo a learning process or remember the responses they gave at the first administration. Statistical techniques are available to overcome this problem, but this may lead to rejection of a very Afine-tuned@ instrument which reflects responsiveness in measure rather than random errors. Wilkin et al [182] and Ghiselli et al [231] label this a 'carry-over' effect, which may change the true score or create preservation effects.

Reproducibility provides a partial answer to the question of validity. Furthermore, reproducibility studies can uncover problems in instrument design, respondent instructions, or quality control. Finally, once the reproducibility of a tool is known, that information can be used to judge between the administrations (for example for monitoring or impact assessment).[80] In addition, knowledge about the reproducibility of an instrument can be used to increase the power of a study.

Some researchers [140, 221] distinguish between data and the database. The structure of the database is the format, or list, or questions used to collect data. The data are the information collected from the subjects, and the database is the total information collected using the structure of the questionnaire. Similarly, dietary studies consist of three components: the input, the data collection questionnaire / tool / instrument, and the compiled data / database.

Using the above as starting point, some researchers reason that a structure (that is a data collection tool) cannot per se be tested for reliability, reproducibility or validity. Consequently, reproducibility of data is separated from reliability and validity of data.[221]

Joachim [221], for example, argues that reproducibility (like reliability and validity) is a logical operator that can be assigned the value of true and false. Since data can be reproduced twice in dietary studies - when the data are deemed to be reliable (that is reflecting what it should) and when the data are deemed unreliable (giving the same, but wrong result) - she claims that there is no correlation between the reproducibility and validity of data. She proposes the following mathematical approach, in which reproducibility is seen as a function of specific variables or groups of variables. In this context, reproducibility is affected by three input variables and is described as follows:

Data collected are f(x1, x2, x3 xn)

Therefore, f(person, place, time)

A study could be deemed reproducible of all these variables are pegged. Any change in the variables will result in a change in the degree of reproducibility. The food intake of a *person* (that is the subject) depends on socio-economic status, age, education, and ethnicity. The geographic location influences food availability and price (that is the *place* variable in the relationship). Seasons, trends and the length of the study, form part of the *time* factor.

Gibson [27] uses the term 'precision' when a dietary assessment technique gives very similar results when used repeatedly in the same situation. The precision is seen as a function of the measurement errors and the true variation in intakes (see introductory sections). In addition, the precision of a particular dietary assessment technique depends on the time frame of the technique, the population group under study, the nutrient of interest, the technique used to quantify foods, and the inter- and intra-subject variability.[27]

For quantification purposes, test-retest reliability can be expressed in terms of a correlation coefficient, the so-called coefficient of stability. Coefficients >0.7 are considered to be a minimum standard of stability.[233]

(ii) Consistency between different users (Inter-rater reliability)

This type of reliability investigates the judgment by a rater, and is being tested when two different raters are required to form an opinion of the same measure. The Kappa coefficient of agreement is used to statistically assess the probability of chance agreement. This coefficient can range from -1 to +1, where -1 indicates agreement worse than chance.[182] Alternatively Leedy [234] states that an agreement of 85% is acceptable and indicates the number of items out of the whole set about which the observers agreed. The major source of error with this method is inconsistency between data collectors, usually due to inadequate training. Training is necessary to ensure that all factors relating to the administration of the tools are kept constant.

In the context of nutritional assessment McCall and Cotton [235] determined the inter-rater reliability by comparing results obtained by a dietitian to those of a nurse. They calculated an agreement confidence interval (above 95%) and Kendall's Coefficient of Concordance as indicators of inter-rater reliability. Low inter-observer variation has been highlighted by Kondrup et al [36] and Jones[185] as very important features of reliability of nutrition screening tools.

(iii) Internal consistency

Internal consistency refers to the extent to which all items in a scale measure the same dimension. Statistically it can be seen as an estimate of homogeneity, during which the extent to which individual items are correlated with each other and with overall scale scores are determined. The Cronbach alpha and Kuder Richardson statistics are examples of appropriate statistical approaches in this context. It must be kept in mind that this analysis would only be relevant to measures containing items relating to one dimension.[182] Using this approach, the instrument is administered to the sample once. The Kuder-Richardson formula is used when items are scored dichotomously, whereas the Cronbach coefficient alpha is used when several answers are possible.[233]

Keller et al [184] recommend that this analysis be performed during pretesting of a tool in a developmental sample (which can be the same as that for the initial validation study). This requires a diverse population of subjects with varying intakes. A Cronbach alpha of 0.7 indicates that all of the items contribute to the discriminating power of the scale and there is good internal reliability. Lower values indicate an excess of nuisance items, or too few items in the scale. Values >7 indicate that there are items measuring the same thing, and that some are thus unnecessary.

McCall and Cotton [235] investigated the internal consistency of a nutritional assessment tool: They established the dimensional structure by submitting data to a correlation matrix and principal component factor analysis with varimax rotation. They reasoned that this approach would answer the question "Does the tool ask the right questions or is there some overlap?" High agreement (overlap) was interpreted as that there was overlap (duplication) of areas covered by the questions, while low internal consistency was associated with a multi-faceted nature of items. A similar approach was followed by Johnson et al.[236]

(iv) Equivalent forms reliability

This fourth type of reliability refers to the extent to which two different versions of the same instrument (for example Form A and Form B of a scholastic aptitude test) yield similar results.[234]

2.4.2.3 Measurement of reliability

There is no universal test of reliability. In addition to the above specific types of reliability and the ways to measure them, Mouton and Marais [232] as well as Ghiselli et al [231] mention the use of parallel forms of a test and the split-half method (comparative parts of a test) as possibilities within (psychometric) test construction. In the case of the latter, this would mean that the items in the scale (which are supposed to measure the same attribute) are (randomly) assigned to two different sets. Each set of items should correspond in the way they classify subjects in the study.[231, 237] Furthermore, it may be necessary to re-establish reliability in differing conditions.[182]

In statistical terms, the most common indices to quantitatively describe degree of reliability of measurement are the standard error of measurement and the reliability coefficient. [231] In respect of the latter, Wilkin et al [182] claim that, in general, the accepted reliability standard is 0.5, since random error will then tend to average out in large samples.

Where statistical techniques are inadequate or inappropriate to numerically check for reliability, standardisation and strict quality control measures during data-collection remain the key requirements for reliability. This means focusing on the four areas in which threats to reliability can emerge (see section "Factors influencing validity and reliability").

2.4.3 Validity

2.4.3.1 Definition and principles

Historically, the most common definition of validity is: "A measurement procedure is valid if it measures what it purports to measure." [238] Within the psychometric and education context, validity, more specifically, refers to the appropriateness of inferences from test scores or other form of assessment.[231] Simply stated, this would mean the following: Given a research question or an aim, how useful (that is valid) are the answers (that is the information) provided by the test score? Thus a valid measuring instrument can be described as measuring what it is supposed to measure and as yielding scores whose differences reflect the true differences of the variable being measured rather than random or constant errors.[183] Beaglehole et al [45] state that a study is valid if its results correspond to the truth with no systematic errors and the random error as small as possible.

All of the above definitions require that the truth be known.[50]

Frongillo [239] has stated that validation is the process of determining whether a method is suitable for providing useful analytical measurement for a given purpose and context. He claims that all of the following criteria must be fulfilled for a method to be called valid for a particular purpose and context:

- Its construction is well-grounded in an understanding of the phenomenon
- Its performance is consistent with that understanding
- It is precise within specified performance standards
- It is dependable within specified performance standards
- It is accurate within specified performance standards
- Its accuracy is attributable to the well-grounded understanding for that purpose and context

More than 40 years ago, Becker et al [238] stated that when the researcher has a (i) perfectly calibrated tool and a (ii) purely objective technique for its use, the error variations tend to be at a minimum. However, nutritionists are faced with a far more difficult problem in accurate data collection:

There is no ultimate criterion-measuring device that can be used for the calibration of other devices (no 'golden standard'), and very often no established criteria exist. Consequently comparative validity, which poses special challenges, judgement and logical and empirical processes [240] is the only alternative, since validating 'usual' or 'habitual' intake presents overwhelming practical difficulties or is actually impossible.[50]

Buzzard and Sievert [32] state that a calibration study involves the collection of dietary data from a subset of study participants by using two different dietary assessment methods. It is the comparison of one method of dietary assessment to another with the aim of better understanding the level of agreement or relationship between the two methods. There is much overlap in the nature of the studies to which the principle is applied. Typically a less-detailed method (test method) is compared to a more detailed method, which is assumed to provide more accurate estimates of intake. The objective of the calibration study is to quantify the bias of the less-detailed method in relation to the more detailed method. This permits adjustment (calibration) at the group level of intakes derived from the less detailed method. Livingstone and Black [159] state that calibration studies are studies of relative validity to distinguish them from studies of validity that use external markers of intake. Given that a gold standard for validation is not available, the best one can do is to calibrate one method against another believed to be more accurate for the purpose at hand.[241] Since validation is hardly possible, calibration is a useful alternative, including comparison of results between studies, between methods and over time. Calibration is the measurement of the distance between measurements from two different instruments, or the measure of change in the accuracy of measure of an instrument over time.[241]

From the above it is evident that the terminology around validation and calibration in the nutrition context has not been standardized.

2.4.3.2 Types of validity

Since the validity of an instrument is a function of the specific aim it is intended to achieve, it follows that different types of validation or validity evidence can be obtained. Thus, questions about validity cannot be separated from a consideration of the specific purpose for which a test is to be used. Equally, a test may have several purposes, which can vary in kind and scope. Consequently a given test may have a moderate validity in achieving one aim, yet have good or poor validity in another respect.[231] Typical examples from the nutritional epidemiology literature emerge where a test might exhibit a different validity in respect of quantitative precision, versus classification agreement, versus ranking of individuals or establishing prevalence in groups.

In clinimetry, which focuses on the quality of clinical measurement, where quality includes both the quality of the measurement instrument and the quality of performance of the actual measurements, [242] this has lead to the convention of referring to different types or forms of validity, each of which is important in different situations.[183, 234, 240]

Face, content, representative and/or consensual validity, as a group, refer to the overall relevance, adequacy and / or (relative) representativeness of the components of an instrument, as judged by content experts and / or potential users of the tool.[36, 233, 234, 240, 242] Sometimes a differentiation between the terms is made, but, generally speaking, these types of validity are related, are based on judgment and tend to be subjective.[240, 242]

Criterion-related validities range from situations where an external 'gold standard' (that is the criterion which reflects 'truth') is available, to relative validity where either the truth is unknown or not (yet) measurable.[183, 193, 231, 240, 242] Establishing these forms of validity tends to be more rigorous than in the above-mentioned group. It has been claimed that these validities (particularly criterion validity) put validation "on the road to good science." [243] An example of this most powerful approach in the nutrition context is the use of doubly labeled water as the criterion for energy expenditure. In the case of relative validity, the measurement obtained by the test method is compared to the results from another method or outcome variable assumed to be more accurate or indicative of the truth. Examples are *construct validity*, where the truth is a trait hypothesised to exist, but there is not one real-world counterpart for it.[122, 183, 231] In this context *convergent validity* refers to the agreement or correlation of independent measurements that are theoretically or logically related.[233, 238, 240, 244] *Discriminant validity*, on the other hand, is inferred when a measurement of a construct successfully discriminates between people known to have differing amounts of the trait being represented by the construct.[233, 244, 245] *Predictive validity* is a type of criterion-related validity that refers to the accuracy with which future outcomes (for example growth in the form of weight for height, weight gain) is forecasted by the test method.[231, 238, 240, 246] A substitute for predictive validity can be *concurrent validity*. In this case, another, currently present trait is measured in the place of the future outcome. From the above, it is clear, that construct validity is not an aspect of validity that is exclusive to other types of validity. In the nutrition literature some of these types of criterion-related validities have been used interchangeably (for example references [198, 208, 235]) or have been applied differently (for example when validating fruit and vegetable intake the use of biomarkers was called criterion validity by Murphy et al [208] and construct validity by Bodner et al (1998).

2.4.3.3 Validity of (dietary) screening tools

A screening tool is valid if it correctly categorises people into groups with and without disease, as measured by its sensitivity and specificity.[45] Caan et al [49] have consequently suggested that sensitivity and specificity are the best indicators of validity of a method for dichotomous classification because of the ability of these parameters to generalise results to populations amongst whom the prevalence of the phenomenon varies markedly. Both, sensitivity and specificity are thus descriptors of the accuracy of a test.

Sensitivity is the proportion of truly ill people in the screened population who are identified by the screening test.[45] The greater the sensitivity of a test, the more likely that the test will detect persons with the condition of interest. Thus, sensitivity is measured in the group of subjects who test positive by the reference method or 'golden standard' and reflects the true positive rate.

Specificity is the proportion of truly healthy people who are identified by the screening test. [45] The greater the specificity, the more likely that persons without the condition of interest will be excluded by the test. In a group of subjects who test negative on the reference method, specificity is defined as the true negative rate in the screening procedure.

A high sensitivity for a screener may give false positives, with more subjects classified as having a high fat intake than is actually the case. Equally, a high screening specificity may give false negatives (based on reference [247]).

Thus having both, high sensitivity and high specificity, represents the ideal. However, a balance must usually be struck between the two, because the cut-point between normal and abnormal is usually arbitrary, and because very often sensitivity and specificity are inversely related. Apart from the inherent aim of the screener, availability of funds and resources to support interventions, the seriousness of the disease, the distribution of the risk factor as well as local experience of the severity of risk, are amongst the factors that will determine whether high specificity or high sensitivity will be favoured in a particular situation. In general, raising a threshold for considering a result to be positive typically will lead to a gain in specificity (fewer false positives) but a loss in sensitivity (more false negatives or missed cases). On the other hand, lowering the threshold for considering a result to be positive typically will reduce the level of false-negatives (raise sensitivity) and increase the likelihood of false-positives (lower specificity). Very specific tests are often used to confirm the presence of a condition.[45]

In conjunction to the above, *predictive value* affects the usefulness of a screener. Predictive values depend on sensitivity and specificity, but most importantly, on the prevalence of the condition in the population tested or the pre-test probability that a subject has the condition of interest. Positive and negative predictive values can be distinguished, where positive predictive value is defined as the percentage of persons with positive test results who actually have the condition of interest, and negative predictive value is the probability of the condition being absent if the test is negative. Thus these two measures address the estimation of probability of disease or a specific condition of interest (for example high fat intake).[45]

Relative risk (risk ratio) is the ratio of the risk of occurrence of a disease among exposed people to that among the unexposed whilst the *odds ratio* is the ratio of the odds of exposure among cases to the odds in favour of exposure among controls.[45]

2.4.4 Validation studies

2.4.4.1 Background

Comparative validation is not new: As early as 1942 Huenemann and Turner published an exemplary validation study in children aged six to 14 years, where, at the beginning, they obtained a detailed diet history from each of the subjects. This was followed by a ten to 14 day precise weighed diet record, which was repeated every three to four months, three or four times, so that the period of time covered for each child ranged from at least six months to one year. Based on the variation of the amount of nutrients, the authors concluded that no single diet record could be considered 'typical' of an individual subject's food intake.[238]

An early review of dietary intake methodologies and validation studies was compiled by Becker et al [238]. In the 1980's various additional reviews on dietary validations were published.[50,77, 248, 249] In the 1990's Friedenreich [91] reviewed methods that measure past diet and Gibson [171] wrote a general review about dietary assessment. Jones [185] specifically critiqued dietary assessment methodology. For the past few years a register for dietary assessment calibration and validation studies has been available online (www-dacv.ims.nci.nih.gov/).

2.4.4.2 Validation studies in children

As mentioned before, Mc Pherson et al [25] published a review of validation studies in school-aged children. The following discussion is thus limited to studies focussing on validation of screeners and studies not discussed previously (under FFQ or food records) or mentioned in the McPherson et al review.

A seven-item fruit and vegetable FFQ had a low validity among third-grade students when compared to seven-day food records.[250] The major problem was the severe overestimation by the FFQ. Cognitive problems were offered as main reason. Field et al [195] compared four brief questionnaires for measuring fruit and vegetable intake with estimates from three 24-hour recalls on non-consecutive days in adolescent. They found the short methods useful for ranking but not for estimating prevalence of consumption of five or more servings of fruit or vegetables per day. The validity and reproducibility of a questionnaire aimed at assessing fruit and vegetable intake was evaluated in sixth grade Norwegian children. It was compared to seven-day food diaries. Reproducibility was acceptable, as was the comparative validity of vegetable intake. Fruit intake, however, was overestimated.[251]

Baranowski et al [252] assessed the validity of a 'Food Intake Recording Software System' against observation of school lunch and a 24-hour recall. They concluded that this lower-cost approach was promising, though somewhat less accurate than the 24-hour recall.

Jonsson and Gummeson [226] assessed reliability and construct validity of a method that utilised picture stacking to measure food choices (milk, margarine, bread, cereals) for breakfast. They reported that for milk and margarine reliability and construct validity were good, but random error or a trend towards healthier choices played a role in the other cases. Pictures were also used for adults in the Cardiovascular Health Study. Food frequency scores were obtained from a picture sort procedure, which yielded relative validity similar to conventional FFQ.[253]

'Yesterday's Food Choices' is a 33-item instrument validated for American Indian children in fifth to seventh grade.[92] A modified diet record-assisted 24-hour recall was validated by direct observation among third-grade American Indian children. Weber et al [254] concluded that at group level the reported macronutrient proportions of total energy intake were accurate.

As part of the Child and Adolescent Trial for Cardiovascular Health (CATCH) a short Food Checklist (CFC) was developed as a measure of PFE, PSFE and sodium intake in middle school students. Children (n=365 seventh graders) provided yes / no responses with respect to intake on the previous day for 40 items on the checklist. Results were compared to 24-hour recalls and reproducibility and validity were demonstrated.[255]

Habitual meal patterns and intake of foods, energy and nutrients in 15-16 year old Swedish girls was measured with a 'diet history'. Seven-day food records served as reference method. The former was found to perform as well as the reference method in terms of classification agreement of meal patterns. Also energy and nutrient intakes were similar. For individual foods there was less similarity.[256]

In Pretoria, South Africa, a modified diet history was compared to seven-day precise weighing food records in six to eleven-year old white children. It was concluded that the shorter and more practical modified diet history gave results at least as satisfactory as the laborious and time-consuming seven-day precise weighing.[63]

Potgieter and Fellingham compared a 24-hour weighing method with a seven-day weighing method in black, Indian and coloured children in Pretoria. They concluded that there was no serious bias in the 24-hour weighing method and that it could be a rough estimate of population means.[173]

2.4.5 Factors influencing validity and reliability

Mouton and Marais [232] identified four major variables to keep in mind when attempting to ensure that validity and reliability are not threatened. Some of these fall in the cognitive perspective, whilst others could be classified as relating to the situational perspective.[257]

- *The researcher / interviewer/field worker* is the first factor mentioned by Mouton and Marais.[232] In this regard the researcher's characteristics such as affiliation, image and distance from the participants, as well as his/her orientations such as bias-producing cognitive factors, attitude structure expectations and role expectations can play a role. Referring to measurement of diet, Gibson [220] specifically mentions the use of incorrect questions, incorrect recording of responses, intentional omissions, biases associated with the interview setting, distractions, confidentiality and anonymity of the respondent, and the degree of rapport between interviewer and the respondent. In the Bogalusa study Frank et al [258] showed that interviewer recording practices had an effect on the recorded nutrient intakes of children.

- The individual who participates in the research project (*participant / respondent / subject*) who, in the so-called guinea-pig effect can show signs of memory decay, omniscience or interview saturation is the second factor highlighted by Mouton and Marais.[232] They add the perceived role, level of motivation and response patterns as being participant orientations that can also influence validity and reliability. Leedy [234] refers to this phenomenon as the reactivity or Hawthorne effect, and states that it specifically is a threat to the internal validity of a study. Diet-related examples given by Gibson [220] in this regard include over-reporting of ‘good’ foods such as fruits and vegetables and under-reporting of ‘bad’ food such as fast foods and alcohol, leading to a so-called prestige bias. Memory lapses, like forgetting to report the ‘minor’ parts of a meal (for example dressings), inability to report portion sizes, and the so-called flat slope syndrome, whereby respondents tend to overestimate low intakes and underestimate high intakes, are further examples.[220] Whilst respondents’ inability to estimate their intake reliably is an important factor influencing reproducibility, Block and Hartman [80] state that methodological explanations are more likely to play a role.

- Mouton and Marais [232] have listed the *measuring tool* (questionnaire / interview schedule) as a third factor which affects validity and reliability. They identified question sequence, open / closed questions, ‘don't know’, mid-position selection, questionnaire length, item sensitivity, leading questions and fictitious attitudes as aspects of importance. The sources of measurement error identified by Gibson, [220] that could be classified under this heading, include coding and computation errors (as when 'standard / reference'quantities of intake are do not reflect the intake of the subject and when these intakes are incorrectly converted to grams eaten). Another source of error can be found in the compilation of nutrient composition data (which can be random, systematic or true [like geographic / seasonal] variability, or due to errors in the nutrient analysis of food items, or the compilation of the computerised data base). An important factor affecting specifically the reproducibility of a tool is the variability it permits.[80] An instrument which does not include portion sizes, or which has limited response categories about frequency of consumption, is likely to have a higher reproducibility score, because it allows less variability. In such a case high level of reproducibility is desirable, but not sufficient to ensure validity. The physical questionnaire design (for example layout) and instructions given to subjects can also affect reproducibility.[80]

- The final factor listed by Mouton and Marais [232] is the *research context* (broad or specific spatio-temporal circumstances). This refers to time, cultural and political factors as well as the research setting as such. Leedy [234] illustrates this threat to internal validity in terms of subject selection, for example the use of volunteers and convenience sampling, and calls it 'experimenter expectancy' which may lead to a selection bias. Within the dietary assessment context, Block and Hartman [80] point out that reproducibility is clearly influenced by the elapsed time between two administrations.

2.4.6 Implications

Validity in nutritional assessment is not a 'black and white' issue: Firstly, because no criterion exists, the focus is on relative or comparative validity and varying degrees of validity are observed. Secondly, validity is dependent on the population and the context. Thirdly, it is important to differentiate between the validity of the measurement instrument and the actual performance of the measurement. If the measurement is performed sub-optimally, the instrument may be sufficiently valid, but the performance may not.[242]

In general, Leedy [234] lists four possible precautions that can help to enhance the internal validity of a study: Controlled laboratory settings, double blind experiments, unobtrusive measures and triangulation. For qualitative research, strategies that can be added include: spending extensive time in the field, performing negative case analysis, obtaining feedback from others and respondent validation. On the other hand, in order to improve the external validity, real-life settings, the use of representative samples and replication in a different context can be considered. Practical and ethical considerations usually necessitate compromise.

As far as reliability is concerned, Leedy [234] stresses consistent administration of instruments. This implies standardisation from one situation or person to the next. Secondly where judgments are required, specific criteria should be established to indicate the kinds of judgments that must be made. Finally any research assistants who are using the tool should be well trained so that they obtain similar results.

In respect of dietary assessment Gibson [220] and Kohlmeier [259] point to the following practical implications:

Quality control needs to be implemented at each stage of the dietary assessment. Quality control refers to the range of procedures undertaken during data collection and analysis to ensure quality of measurement. This involves steps to prevent, reduce, detect and correct errors. The following aspects deserve special attention:

- Researcher and field workers:
  - Training and retraining for interviewers and coders, referring to aspects such as the extent of probing and use of probing aids, wording of questions, participation of other persons in data collection [260]
  - Standardisation of interviewing techniques and questionnaires
  - Pre-testing of questionnaires
  - Pilot surveys
  - Training of interviewers to anticipate and recognise potential sources of distortion and bias
  - Minimise non-response by training interviewers to convey understanding, trust and warmth
  - Concentrate at avoiding value judgments
- Respondent:
  - Implement knowledge about cognitive processes involved in diet recall, specifically to improve question comprehension, improving information retrieval, improving estimation of quantities, improving response formulation [91]
  - Attention to memory by using probes, visual aids et cetera
- Data handling and computer program:
  - Credibility of software and nutrient database

    - Reduction of number of steps in data processing
    - Duplicate entry for a certain percentage of observations
    - Programming for error detection (frequency distributions, flagging et cetera) [260]
    - 'Coding rules' to deal with incomplete or ambiguous food descriptions / meal codes
- Sampling:
- Collection of supplementary information

In conclusion, from the review of the literature it is evident that the dietary habits of children are very often not in line with international recommendations, particularly in respect of fat intake and for reducing risk of CNCD. Dietary assessment of children can take on many forms, but no one method is perfect. Screening appears to be an attractive alternative, but, as in the case with detailed assessments, the comparative validity of these methods must be established in the population for which the tool is intended.

CHAPTER 3:DEVELOPMENT AND DEVELOPMENTAL EVALUATION SUB-STUDIES OF RESEARCH TOOLS

Chapter overview

3.1 TEST METHOD
3.1.1 Development rationale, aims and correct application information
3.1.2 Format
3.1.2.1 Item list
3.1.2.1.1 Background (original tool)
3.1.2.1.2 Application (current tool)
3.1.2.1.3 Sub-study 1: Content and face validity
3.1.2.1.4 Practical development
3.1.2.2 Quantification
3.1.2.2.1 Background (original tool)
3.1.2.2.2 Application (current tool)
Sub-study 2: Reference portion size
Sub-study 3: Portion size estimation aids
Sub-study 4: Frequency of intake
3.1.2.2.3 Scoring

3.2 REFERENCE METHODS
3.2.1 Reference method 1: Food record
3.2.1.1 Background
3.2.1.2 Sub-study 5: Food record
3.2.1.2.1 Form
3.2.1.2.2 Time frame and programme
3.2.1.2.3 Sample
3.2.1.2.4 Training and briefing
3.2.1.2.5 Findings
3.2.2 Reference method 2: Screener by parents

3.3 REFLECTION

3.1 TEST METHOD

Keller et al [184] and Jones [185] recommend that the first step in the systematic development of a nutritional assessment and screening tool should be to spell out the aim(s) of the instrument. Only when that, which is to be determined by the screening tool, is very clear, can the validity thereof be measured. Furthermore, since the context can also affect validity, the correct usage (for example the relevant outcome variable[s], target group and administration detail) should be specified.

3.1.1 Overall development rationale, aims and correct application information

Based on the possible purposes and uses that Wilkin et al [182] identified for health-related measures, the current screening tool should have primarily discriminative properties, since it should separate children with a high fat intake from those eating according to the prudent dietary guidelines as specified by the NCEP.

Within clinical practice, the dietary fat screener should enable health professionals to correctly classify individuals according to their fat intake. Since the valid measurement of usual, absolute dietary intake of individuals appears to be an unrealistic aim, the correct ranking of groups of children is the immediate aim, primarily for use in nutritional epidemiology. Thus, usual dietary intake of PFE, PSFE and cholesterol should be reflected by the dietary screener, even though quantitative precision in terms of these variables was not required.

Since the screener was supposed to be a tool for level 1 [177] nutrition care, the current tool had to be administrable by qualified professionals (that is 'individuals who are qualified by virtue of their education, experience, competence, applicable professional licensure, regulations, or certification, registration, or privileges') who have a basic nutrition knowledge and who have knowledge of and access to referral systems for children who are identified as having a nutritional risk. Alternatively, it should be suited for self-completion. No parental assistance should be required for providing the information. The setting was school-based requiring no specialised apparatus. The target group was urban, middle-class South African school children in grade six.

The following practical usage and development criteria were set:

- South-African food-based so as to provide a food-based starting point for behaviour change counselling in line with current thinking within nutrition education theory
- Requiring minimal literacy from participant

- Attractive: Pictorial, colourful, novel and practical
- Interpretable and action-oriented
- Administrable in group setting
- Suitable for school settings
- Quick to use

Following the developmental evaluation and comparative validation (thus prior to implementation), the scientific merits of the screener should be known in quantitative terms.

3.1.2 Format

The basic structure chosen for the test method was that of a FFQ, since a FFQ can accommodate most of the stated aims and pre-set criteria, for example estimation of usual intake, focusing on specific dietary components with relatively low respondent burden. Consequently the general guidelines for developing such a tool were followed [26, 27, 115] and in particular the principles regarding cultural appropriateness.[116, 261]

In the absence of representative food intake data for the target group, the US's NCEP MEDFICTS Dietary Assessment Questionnaire was chosen as starting point for the development. The MEDFICTS tool is widely published in standard nutrition text books (for example references [10, 26]) as well as on the world wide web (www.bgsm.edu/nutrition/medfic.htm; www.nih.gov/news/stepbystep/medficts.htm; http://www.nhlbi.nih.gov/guidelines/cholesterol/atp3full.pdf).

Since food questionnaires should be adjusted for the population group for which they are intended,[116] a number of changes were made to the original MEDFICTS tool during the initial development for this study. In the following sections the item list and the various elements of the quantification are each described in terms of background considerations and the application in the current tool. In addition, the developmental evaluation sub-studies that were conducted in terms of the above are briefly presented. The Research and Ethics Committee of the University of Pretoria approved each of the developmental evaluation sub-studies individually.

3.1.2.1 Item list

Creating an item list is the first step in the development of a FFQ type dietary assessment tool. This refers to the foods or rows in the questionnaire.

3.1.2.1.1 Background (original tool)

In the original MEDFICTS tool, the individual food item list is replaced by eight food categories (Meats, Eggs, Dairy, Fried foods, In baked goods, Convenience foods, Table fats, Snacks), which are the prime contributors of dietary fat, saturated fat and cholesterol in the North American diet.[201] Kris-Etherton et al [203] created three sub-categories (milk, cheese and dessert) for the dairy category. Furthermore, each category has a group 1 (high fat choices) and a group 2 (low fat choices).

3.1.2.1.2 Application (current tool)

A consensus workshop consisting of three private practicing dietitians (with extensive experience in dealing with the target group) and the researcher reviewed the original item list for face and content validity. It was decided to retain the main MEDFICTS categories and sub-categories. However, in the meats, eggs, fried foods, in baked goods, and convenience foods categories group 2 was excluded, since it was reasoned that in the South African context this would cause practical problems because of limited availability, access and target group awareness of these foods. Furthermore, including this group did not affect the score. The role of the group 2 foods in the original tool was only for clarification purposes.[203] Thus a total of ten food categories formed the basis of the scored part of the current tool with five clarifying food categories (see Table 3.2 for a text summary of the tool).

Subsequently a developmental evaluation sub-study was conducted to obtain a more objective indication of the content and face validity.

3.1.2.1.3 Sub-study 1: Content and face validity

Rationale: The face validity of a method describes, according to Johnson et al,[262] the extent to which the questions asked conform to current expert opinion relating to what the instrument is intending to measure. It is based on intuitive judgment of experts, and is considered a necessary step in measuring the validity of any new dietary tool.

Aim: To identify the food items recommended by local dietitians to healthy black and white children between the ages of nine to twelve years in order to meet selected food-based dietary guidelines.

Participants: Anonymous questionnaires were handed out to 120 dietitians attending Association of Dietetics in Southern Africa meetings in the period June to August 2001 in the branches Mpumalanga, Pretoria and Gauteng South (South Africa).

Methods: Dietitians were requested to list the ten most important foods for children in the specified age group to increase and to decrease in order to achieve selected food-based dietary guidelines. During data analysis final year B Dietetics students, who were unaware of this sub-study's aim, grouped the listed foods. The groups that were formed were: '*fatty meats'* consisting of regular minced meat, steak, bacon, biltong [that is dried beef], dried beef sausage, boerewors [that is beef sausage], lamb chops and sausages; '*processed meat*' included russions, polony, viennas, salami and corned beef; '*organ meat'* included tripe, liver and kidney; full cream cheeses, full cream yogurt, ice cream, cream, full cream inkomazi [that is soured milk] and milkshakes formed the '*full cream dairy product'* group; '*take-aways'* referred to hamburgers, toasted sandwiches and pizzas; '*animal fat*' included lard, Holsum [that is hydrogenated fat] and butter.

Results and discussion: A total of 50 questionnaires were timeously returned by prepaid postage. Ten of these had to be excluded because they did not meet the inclusion criterion of coming from dietitians who consult the target group at least once a week, resulting in a final sample size of 40. In Table 3.1 the findings regarding the ten most frequently mentioned foods / groups of foods to eat less of in order to lower fat intake are summarised. Seventeen and 23 dietitians reported that their respective majority clients were white and black children. They thus based their recommendations on that group. (Results regarding foods to increase in order to lower fat intake, and recommendations in respect of the other guidelines are not reported).

TABLE 3.1: FOODS TO EAT LESS OF FOR LOWERING FAT INTAKE: NUMBER (PERCENTAGE) OF DIETITIANS (n=40) WHO MENTIONED EACH FOOD / FOOD GROUP

White children (n=17)		Black children (n=23)		All children (n=40)	
Rank and food	**Frequency (%)**	**Rank and food**	**Frequency (%)**	**Rank and food**	**Frequency (%)**
1 Chocolate	14 (82)	1 Full cream dairy products	20 (87)	1 Full cream dairy products	33 (83)
2 Full cream dairy products	13 (76)	2 Fried chips	19 (83)	2 Fried chips	31 (78)
3 Fatty meats	12 (71)	3 Processed meats	14 (61)	3 Fatty meats	24 (60)
4 Fried chips	12 (71)	4 Fatty meats	12 (52)	3 Chocolate	24 (60)
5 Biscuits	9 (53)	5 Chocolate	10 (43)	5 Processed meats	19 (48)
5 Take-aways	9 (53)	6 Margarine	9 (39)	6 Margarine	17 (43)
5 Margarine	9 (53)	7 Crisps	9 (39)	7 Crisps	16 (40)
8 Crisps	7 (41)	8 Fish	8 (35)	8 Biscuits	14 (35)
8 Animal fat	7 (41)	9 Organ meats	8 (35)	8 Take-aways	14 (35)
10 Chicken	6 (35)	10 Cakes	7 (30)	8 Animal fats	14 (35)
		10 Coffee / tea creamers	7 (30)		
		10 Chicken	7 (30)		
		10 Animal fat	7 (30)		

From the survey the following circumstantial evidence emerged:

- In general, the foods listed showed some similarity with the MEDFICTS food categories, even though the groupings differed. Eggs, regardless of racial group, were a clear exception in the sense that they did not feature at all in the top ten listings.
- There tended to be more agreement between items in the test method (MEDFICTS) and the recommendations for white children, than between the test method and recommendations for black children. This confirmed that a valid item list is population-dependent.[111]
- In the top ten sources of fat in the diets of twelve-year-old Australian children Gracey [263] reported a list very similar to the MEDFICTS foods and the food listed in this sub-study, except for eggs which do not appear on that list.
- The list also showed some similarity with the findings of Johnson et al [262] except that meats, eggs and convenience foods did not feature on their top ten list for British school children, again emphasising the importance of contextualisation.

It was thus concluded that the test method had reasonable face validity for the current target group of white South African children. Further support for face and content validity comes from the fact that this list showed substantial similarity with numerous published fat screeners, for example Caan et al [49], who modified Block et al's original fat screener [206]; the 'Dietary Risk Assessment ' [213, 214]; the 20-item short questionnaire to qualitatively assess the intake of total fat, saturated, mono-, polyunsaturated fatty acids and cholesterol of Rohrmann & Klein [264]; the brief FFQ for fat, fiber, and fruit and vegetables intake of rural adolescents by Buzzard et al. [265] In addition, recent analyses of the dietary sources of nutrients among a nationally representative sample of US adults [22, 266] yielded on visual inspection similar results in terms of the food sources of fat, saturated fat and cholesterol.

3.1.2.1.4 Practical development

For the current tool, colour pictures of South African (branded and generic) foods within each food category were used instead of text for the following reasons:

- Jonsson et al [267] have argued that the reliability, validity and usability of a measure of dietary habits can be increased if the measure does not require sophisticated linguistic abilities, and concentrates more on visual and comprehension skills.
- In a society with many languages this has obvious additional advantages. An instrument of this nature is presumably more likely to be suitable for a trans- or multicultural target group.
- This approach would make the tool also suitable for children with limited literacy.
- Visual appeal can never be wrong, even though it may be possible that brands could also cause problems, for example amongst brand loyal consumers whose particular preference is not pictured, even though it is implied to be part of the category because of similar nutritive value. Equally, the depicted example food may be a particular brand not chosen by a respondent, or the availability of certain brands (for example yoghurt) may be regional. Finally, new foods tend to enter the market at a very fast rate.

In order to minimise the limitations, yet keep the advantages of package recognition, a clarifying introduction (in standardised text) was always offered when a new category was presented to the respondent. This specified what the picture under discussion showed, for example, in the case of meats the introductory statement was as follows: "This is a picture of various sorts of meat. It ranges from beef like steak, boerewors, biltong and minced meat, to pork, mutton and chicken. Organ meat, like kidneys and liver, is also included. Processed meats, such as cold meat, bacon, polony and spreads also form part of this group." Table 3.2 provides an overview of the item list (food categories and pictured example foods) and in Addendum A are reduced copies of the pictures in the tool.

For the current study, the first question on the initial tool was always: "Do you eat foods such as those on the picture?" For those participants who responded positively to this filter question a follow-up question "Which one food do you eat most of the time?" was posed, because Koehler et al [92] suggested that children more accurately recall specific items rather than categories in their ‘Yesterday’s Food Choices’ instrument.

TABLE 3.2: TEST METHOD: DESCRIPTION OF FOOD CATEGORIES, FOOD PICTURES INCLUDED, REFERENCE PORTION SIZE AND PORTION SIZE ESTIMATION AIDS (PSEA)

Category	**Sub-category / group**	**Code**	**Food pictures**	**Reference portion size and PSEA's (P=photo), (D=diagram) (H=Household measure)**
Meats	-	M	Beef: steak, boerewors, biltong, minced meat Pork: Chicken: Mutton: chops, roast Processed meat: cold meats, bacon, polony, spreads Organ meat: kidney, chicken liver	90mm diameter circle (D)
Eggs	-	E	Eggs	2 eggs (P)
Dairy	Milk, whole or fat reduced	DM1	Fresh or long life milk (full cream or 2%) Milk powder (full cream or blends) Coffee/tea creamers Condensed or evaporated milk Full cream yoghurt	250mL (H)
	Milk, skim/fat free	DM2	Fresh or long life fat free milk Powdered fat free milk Fat free / low fat yoghurt or drinking yoghurt	250mL (H))
	Cheese, full cream	DC1	Hard cheese (full cream) Processed; spread or wedges (full cream) Cottage/ cream cheese (full cream)	3 slices ea 30x90x2mm (D) 2 wedges (D) 125mL (H)
	Cheese, fat reduced	DC2	Hard, low fat cheese Processed low fat; spread or wedges Cottage cheese, low fat	3 slices ea 30x90x2mm (D) 2 wedges (D) 125mL (H)
	Dessert, full cream	DD1	Milkshakes Full cream ice cream Full cream custard Full cream dessert (blancmange type)	125mL (H)
	Dessert, fat reduced	DD2	Low fat flavoured milk drinks Low fat frozen dessert Frozen yoghurt	125mL (H)
Fried foods	-	F	Chips (French fries) Fried vegetables (eg onion rings) Fried chicken (eg whole or pieces) Fried fish (eg hake) or fried seafood (eg calamari) Fried meat (eg sausage) or fried eggs	90mm diameter circle (D) 125mL (H)
In baked goods	-	I	Cakes Cookies/ biscuits Sweet tarts / pastries (eg chelsea buns, doughnuts, eclairs) Savouries (eg samoosas, croissants, vetkoek) Rusks	120mm diameter circle (D)
Convenience foods	-	C	Tins / cans (eg spaghetti) Packaged (eg pasta sauces, noodles) Frozen meals (eg pizza)	100x100mm square (D) 250mL (H)
Table fats	High fat	T1	Butter Margarine brick wrapped in paper Mayonnaise / salad dressing Peanut butter	5mL (H)
	Fat reduced	T2	Low fat salad dressing / mayonnaise 'Lite' / medium fat tub margarine	5mL (H)
Snacks	High fat	S1	Chocolates Chips and cheese puffs (eg Niknaks) Regular savoury crackers Peanuts	Photo: 1 small packet chips 50g chocolate 22g peanuts 6 crackers, high fat
	Fat reduced	S2	Low fat crackers or pretzels Fruit bars Hard candy, liquorice, jelly sweets, marshmallows	3 ProVita (P) 1 fruit bar (P) 10 sweets, low fat (P) 30g pretzels, plain (P)

3.1.2.2 Quantification

3.1.2.2.1 Background (original tool)

According to Block et al [44] quantification includes the portion or serving size attributions of each line item and also the assumed nutrient content of each. In addition, the measurement of frequency of intake should also be considered since all of these eventually determine the relevance of the score obtained.

3.1.2.2.2 Application (current tool)

In the current tool the following three specific aspects of the quantification were investigated in more depth in the proposed target group: the reference portion size, the portion size estimation aids (PSEA) and the frequency of intake.

- ***Reference portion size***

Most of the reference portions in the original MEDFICTS tool are consistent with definitions given by the American Diabetes Association exchanges, American Heart Association, and the US Department of Agriculture handbooks.[203]

These original reference portion sizes were adjusted for most food categories since it was felt that the relevance for the target group was in some cases questionable. The same team of dietitians mentioned before, again held a consensus workshop. Recommendations were made based on their experience and in some by cases practically weighing and measuring example foods. In order to field-test these reference portion sizes a small survey was conducted.

For the purpose of this study small portion was defined as half as much or less than the reference portion. A large portion was equivalent to one-and-a-half times or more the size of the reference portion. In the original tool 'less' or 'more' than the reference were seen as small and large respectively.

Sub-study 2: Reference portion size

Rationale: Ease of completion and brevity are desirable for screeners. Quantitative FFQ's (such as MEDFICTS) require that respondents specify their usual intake relative to a given medium portion. The question arose whether the proposed medium portion size was valid for the intended target group, and whether the screener could be simplified to a semi-quantitative format (in which medium intake was assumed).

Aims: To explore ten to twelve year old children's perception of a medium portion and the actual amounts dished up by them within the context of a given meal for three different example foods from the test method.

Participants: Fifty-two children (21 male and 31 female), who met the inclusion criteria of being familiar with and not allergic to the test foods were systematically chosen from the grade four, five and six class lists of an urban primary school.

Methods: Each child was requested to dish up the amount of test food (minced meat, chips and margarine) that was usually eaten. For minced meat the plate already contained cooked white rice and carrots. The chips had to be added to a fast food hamburger, and the margarine had to be spread onto a 40g slice of bread. The amount dished up was weighed on an electronic Soehnle kitchen scale. For assessing the perception of medium size portion, four different portion sizes of each test food were then presented in random order to each participant on separate plates: 50%, 100%, 150% and 200% of the proposed reference portions of the test foods, that is 90g minced meat, 45g chips and 5g margarine, keeping the accompanying foods identical. After an introductory definition ("A medium portion is the amount eaten by children to stay healthy"), the participant had to report which amount, in his/her opinion, reflected a medium portion. Test-retest reproducibility was checked by having each fifth (10%) subject repeat the assessment.

Results and discussion: The mean age of the participants was 10.8±0.8 years and mean BMI was 17.9±2.42kg/m^2. The 45g and 90g portions of minced meat were perceived by about equal numbers (23 and 25 respectively) of participants as "the amount a child eats to stay healthy", even though the mean amount actually dished up was about 45g (Table 3.3). The mean may to be age-related, particularly amongst the boys. The fact that almost 35% of children reported disliking minced meat (data not shown) may also be of relevance. Most children perceived the 40g portion of chips as the medium portion. This was in agreement with the mean amount dished up (44.6±18.92g). No age-related pattern could be detected, but girls always dished up less than the boys (Table 3.3). In the case of margarine, the 5g portion was perceived by the majority of children as 'medium', whilst the mean amount spread onto bread was 4.12±3.76g. Similar to minced meat, many reported disliking margarine (data not shown), and age may also here be a determinant, specifically in boys (Table 3.3).

Conclusion: Perceived medium portions were sometimes (for example for chips and margarine) perceived to be the same as the adjusted reference. Mean actual intake might also be similar (for example for margarine), but inter-individual variations (as evident from the large standards deviations) and possible age and gender-related differences, the small sample size and the limited number of example foods did not yet support the use of one assumed reference portion size per food category as in a semi-quantitative FFQ.

TABLE 3.3: PERCEIVED MEDIUM PORTION SIZE AND MEAN AMOUNT DISHED UP OF MINCED MEAT, POTATO CHIPS AND MARGARINE BY AGE AND GENDER (n=52)

Food	Perceived medium portion size (g)	Number of participants						Total
		10 years Total = 22		11 years Total = 19		12 years Total = 11		
		Male *Total=8*	Female *Total=14*	Male *Total=9*	Female *Total=10*	Male *Total=4*	Female *Total=7*	
Minced meat	45	3	9	1	8	0	2	23
	90	5	4	6	2	3	5	25
	135	0	0	1	0	1	0	2
	180	0	1	1	0	0	0	2
	Mean amount dished up (g)	36.3±12.7	36.5±9.7	59.9±26.2	37.2±9.9	75.0±18.1	44.7±11.3	44.7±18.9 g
		36.4±10.6		48.0±22.1		55.7±20.2		
Potato chips	20	0	4	2	1	0	1	8
	40	5	9	3	7	0	3	27
	60	2	0	3	1	3	3	12
	80	1	1	1	1	1	0	5
	Mean amount dished up (g)	46.9±16.4	32.1±8.8	68.3±23.6	39.3±10.7	59.3±12.8	35.7±8.1	44.6±18.9 g
		37.5±13.8		53.1±23.0		44.3±15.2		
Margarine	2.5	1	5	3	3	0	0	12
	5	4	7	4	6	2	6	29
	7.5	3	2	0	1	2	1	9
	10	0	0	2	0	0	0	2
	Mean spread on bread (g)	3.4±2.4	2.4±2.5	5.9±3.1	2.8±2.3	9.5±8.9	4.9±2.3	4.1±3.8
		2.8±2.4		4.3±3.1		6.6±5.7		

- ***Portion size estimation aids (PSEA)***

The original tool does not include PSEA's, but Kris-Etherton et al [203] do mention that the use thereof could enhance the value of the tool. Two-dimensional serving size measurement aids (life-size photos and geometric shapes) [26, 129] were developed for each food category to assist in estimating quantity of intake. Again, these were the result of a consensus workshop with the mentioned three private practicing dietitians, whose clientele included many children, and the researcher.

In order to field-test the PSEA a developmental evaluation sub-study was conducted:

Sub-study 3: Portion size estimation aids (PSEA)

Rationale: Cost and space favour the use of two-dimensional (2D) PSEA in dietary screeners, but this should not be at the expense of validity and reliability. In adults 2D and three-dimensional (3D) PSEA's appear to yield similar results,[129] but this has not yet been demonstrated in children.

Aims: To investigate the ability of twelve-year old children to express real servings of example foods in terms of 2D and 3D PSEA's, and to compare 2D to 3D PSEA's.

Participants: Of 60 twelve-year old children randomly chosen from the class lists, 22 boys and 21 girls (total 43) and their parents provided timely written, informed consent for participation.

Methods: In an one-to-one, standardised encounter each child was shown a known amount of five true, ready to consume foods (milk, chicken, bread, butter and chips), and requested to indicate for each food which one of three different sized /graded 2D graphics (drawings of measuring cups and teaspoons, cutout circle shapes and rectangles) most closely reflected the quantity of the true food. Similarly, the child was asked to describe the true food as 'half as much', 'the same' or 'one and a half times as much' as a 3D food model (Nasko-Ford Atkinson, WI) of the respective food. Milk and chicken information was obtained twice to control for guessing and check for test-retest reliability. Error rate, defined as the total number of wrong answers expressed as a percentage of total respondents, was calculated for each food and PSEA. The McNemar test for symmetry was used to assess differences / bias.

Results and discussion: In the case of milk, no significant difference between the two subsequent administrations for both the 2D (P=0.32) and the 3D (P=0.48) PSEA was found, thus suggesting test-retest reliability. For chicken, however, the two administrations significantly differed for both PSEA's (P=0.0016 for 3D and P=0.0000 for 2D, based on the fact that in both administrations all children answered wrongly). Thus the guess factor could not be ruled out. From Table 3.4 it is evident that error rates are lowest for milk (for 2D and 3D estimations of quantities) and highest for chicken. The mean error rates for all 2D and 3D estimations (milk 1, butter, bread, chips, chicken 1) were 36.26% and 54.42% respectively. For milk, butter and chips the difference between the 2D and 3D PSEA was not statistically significant (P>0.05). When comparing the error rate in the quantification of milk to foods of differing consistency (that is all other foods) a statistically significant difference (P<0.05) was found for all foods when using 3D PSEA, as well as for butter with the 2D PSEA (data not shown). When checking for gender differences, it appeared that only in the case of chips, using the 2D PSEA, the difference between boys and girls was statistically significant (P=0.01).

TABLE 3.4: ERROR RATES FOR 2D AND 3D PSEA[a], AND SIGNIFICANCE OF DIFFERENCE (P) BETWEEN THEM FOR VARIOUS FOODS

Food	Error rate (%)		P-Value[c]
	2D PSEA	3D PSEA	
Milk[b]	11.6	7.0	0.48
Chicken (1st administration)	100.0	72.1	NA
Chicken (2nd administration)	100.0	95.3	NA
Bread	9.3	51.2	0.03
Butter	34.9	95.3	0.65
Chips	25.6	34.9	0.91

[a] Two dimensional and three dimensional portion size estimation aids
[b] First administration only, because of consistency of response (see text)
[c] McNemar test for symmetry

Conclusions: Inconsistent responses and error rates confirmed that children had problems with estimating quantities of individual foods. The magnitude of the error varied across foods. Physical food consistency, but not gender, might have played a role. However, in line with findings of Cypel et al [129] for adults, 2D and 3D PSEA's did not appear to differ significantly. It is concluded that 2D PSEA's could be retained in the screener, but a reference method in a validation study should include weighing (in contrast to estimation) in order to describe the likely quantification error.

- ***Frequency of intake***

It has been claimed that the main determinant of variation in measuring dietary intakes is frequency of consumption of the food items in the list.[140] In the original MEDFICTS tool three response options are given for categorizing frequency of intake per week: Once or less, up to three times, four or more times. Kris-Etherton et al [203] assert that this frequency grouping would minimise the tendency of persons to underestimate intake, and that it resembles those used in existing questionnaires.

From the review of literature it was evident that reporting frequency of intake is a cognitively challenging task. It was not known whether children would be able to perform these tasks with the necessary accuracy. Another question that arose was whether a (graphic) depiction of frequency of intake, for example as bar charts or abacus type presentations could enhance responses. Whilst graphics might have eye-appeal, they might also be an abstraction, and might reduce rather than improve accuracy in the target group.

A developmental evaluation sub-study was conducted to assess the above-mentioned problems.

Sub-study 4: Frequency of intake

Rationale: The cognitive tasks involved in providing a correct frequency of intake response on a grouped weekly consumption FFQ, include at least the following steps:

1. Recognizing the individual foods / items on the list
2. Understanding the food categorisation (grouping) principle (which foods form part of a specific food category)
3. Recalling own intake
4. Counting number of times per week foods are eaten
5. Summation of counts of foods belonging to a food category
6. Conversion of summated intake to the frequency of intake format used in the data collection tool (recording the answer in the required questionnaire format)

Steps 4 and 6 were the focus of this developmental evaluation sub-study.

Aim: To assess the ability to correctly count and record frequency of food intake in three example cases (reflecting low, medium and high weekly consumption), as well as preferred depiction of this frequency of intake (MEDFICTS response categories versus graphics and absolute numbers), in ten to twelve year old school children.

Participants: A systematic sample of 39 ten to twelve year old children (19 male, 20 female) was selected.

Methods: On an individual, random rotation principle, each child was exposed to three example cases reflecting low, medium and high weekly consumption (corresponding to MEDFICTS grouping). They then had to orally indicate "How many times in a week did [the example case] eat [example food]". This was noted by the field worker. The child then had to record his / her answer on each of four different response formats (MEDFICTS grouping; table format with short sentences and numbers; horizontal abacus-type format; vertical bar-graph format with numbers underneath). Finally the four response formats had to be ranked in order of preference.

Results and discussion: Table 3.5 summarises the findings. As expected, the low frequency of intake case caused no problems either in terms of providing a correct answer or in terms of recording the answer into any of the four given response formats.
For both, the medium and high frequency of intake scenarios, however, five subjects (12.5%) provided wrong answers to the question, two of which overlapped between the two intake

scenarios. This was in line with the findings of Hammond et al [112] who found that food items eaten frequently were, on average, least accurately classified compared with those eaten with intermediate frequency. An inability to give a correct answer and an inability to correctly code the own answer will result in an invalid response. Error rate (defined as the percentage responses that were not correctly answered and correctly coded for each scenario and response format) ranged from 0% to 17.9%, with the majority of error within each of the error rates contributed by wrong answers, in contrast to recording errors. In terms of preference, the words with numbers format was chosen as first or second choice by a total of 32 (82.1%) of subjects, compared to 29 (74.4%) who chose the bar diagram with numbers in first or second place. What these two formats have in common was the provision of numbers. The abacus format and the MEDFICTS grouping were chosen by eleven and six subjects respectively as first or second choice in terms of preferred response format. Apart from a horizontal versus a vertical alignment respectively, the abacus and bar diagram were very similar, except that the bar diagram also provided the corresponding numbers below each bar. It appears that graphics represent an abstract representation or distraction that is disliked rather than assisting correct response. Furthermore, response categories (pre-set groupings) also require an additional step of placing an answer in a suitable group (category), again rather contributing to dislike than making the recording process easier.

Conclusions and recommendations: Children do make reporting mistakes when required to indicate weekly consumption of individual foods, particularly in medium to high intake situations. More of the error was attributable to giving wrong answers in the first place, than to incorrectly recording an answer in a particular response format. Thus, the focus should be on helping the children to correctly count their weekly consumption. However, a more preferred recording format would include the use of concrete numbers (not groupings) and avoiding abstractions (for example graphics). Avoiding groupings would have the additional advantage of eliminating the overlap in the MEDFICTS consumption categories ‘rarely or never’ with ‘3 times or less’. Since summation problems may have made it difficult for the children to calculate weekly intake in the high frequency scenarios, a ‘per day’ response option could be considered for foods with an expected high frequency of intake.

TABLE 3.5: CORRECTNESS OF ANSWERS AND RECORDING FOR THREE FREQUENCY OF INTAKE SCENARIOS, AS WELL AS PREFERRED WAY OF RECORDING OF 10-12 YEAR OLD CHILDREN (n=39)

Scenarios	Answers	Recording							
		MEDFICTS Categories		Words with numbers		Abacus		Bars with numbers	
		Correct *n(%)*	Wrong *n(%)*	Correct *n(%)*	Wrong *n(%)*	Correct *n(%)*	Wrong *n(%)*	Correct *n(%)*	Wrong *n(%)*
High frequency intake	Correct (n=34)	32(82.1)	2(5.1)	34(87.2)	0(0)	34(87.2)	0(0)	33(84.6)	1(2.5)
	Wrong (n=5)	4(10.3)	1(2.5)	5(12.8)	0(0)	5(12.8)	0(0)	5(12.8)	0(0)
	Error rate[a]	17.9%		12.5%		12.5%		15.4%	
Medium frequency intake	Correct (n=34)	34(87.2)	0(0)	34(87.2)	0(0)	32(82.1)	2(5.1)	34(87.2)	0(0.0
	Wrong (n=5)	5(12.8)	0(0)	5(12.8)	0(0)	5(12.8)	0(0.0)	5(12.8)	0(0.0)
	Error rate[a]	12.5%		12.5%		17.9%		12.5%	
Low frequency intake	Correct (n=39)	39(100.0)	0(100.0)	39(100.0)	0(100.0)	39(100.0)	0(0.0)	39(100.0)	0(0.0)
	Wrong (n=0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
	Error rate[a]	0.0%		0.0%		0.0%		0.0%	
First choice		2(5.1)		14(35.9)		5(12.8)		18(46.2)	
Second choice		4(10.3)		18(46.2)		6(15.4)		11(28.2)	
Third choice		6(15.4)		5(12.8)		20(51.3)		8(20.5)	
Least preferred		27(69.2)		2(5.1)		8(20.5)		2(5.1)	

[a]Error rate = ((Total - (Number correctly answered and recorded) / Total) expressed as percentage; that is (39 minus (number correctly answered AND correctly recorded) divided by 39 and multiplied by 100)

3.1.2.2.3 Scoring

Portion size: In the original tool small, average and large portion sizes (relative to the reference) are respectively scored 1, 2 and 3 for the high fat group. The same scoring was used for the current tool. In the original tool no points are given for low fat choices, except if in the meat group 2 portion size was large. In this case 6 points are given. Since no meat group 2 was included in the current tool, these points were of no relevance.

Weekly consumption: In the original tool intakes of three or less servings per week receive three points, whereas a weekly consumption of four or more servings gets rated seven points. This applies to the high fat groups (group 1) of each food (sub)category. No points are given for intakes in group 2 (the low fat group) of the food categories. This same scoring system was used in the present study, based on the rationale given by Kris-Etherton: [203] Foods eaten rarely or never were assumed to contribute an insignificant amount of fat to the diet. Points for weekly consumption were derived from the assumption that a person consuming four or more servings per week (maximum seven) would average 5.5 servings per week. Similarly, persons consuming three or fewer servings per week would average two servings per week.

Four or more servings per week contributed approximately 73% (5.5/[5.5+2]) of the intake for this category of foods and three or fewer servings per week contributed approximately 27%. The simplified multiplication factors under weekly consumption became seven and three respectively.

Category score and final score: For each (sub) category the weekly consumption point was multiplied by the serving size point to obtain a category score. The category scores were added to yield a final score that could thus range from 0 to 216. In the original instrument a final score of 40 to 70 reflects a Step 1 diet, whereas a final score of less than 40 suggests a Step 2 diet. In the Revision 2000 of the American Heart Association the 'Step 1' designation has been replaced by 'major guidelines for the general population' and the 'Step 2' by 'medical nutrition therapy'. As the tool was intended for community-based use, specifically for children, it was decided to only dichotomise the final score into 'high fat' and 'prudent' intake (that is the major guideline), thus omitting the medical nutrition therapy.

3.2 REFERENCE METHODS

As indicated previously two reference methods were used in this study: The food record and parental completion of the screener.

3.2.1 Reference method 1: Food record

3.2.1.1 Background

The three-day food record was the primary reference method in this study. It was therefore considered essential to conduct field-testing in order to optimise the data collection.

3.2.1.2 Sub-study 5: Food record

The completion of the three-day food record was intended to be a mathematics assignment in the new South African outcome-based education approach (Curriculum 2005).[108] The school identified for testing agreed to participate, but requested that all the children in the grade be included. No incentives were given, however, the data collection and recording tasks were considered a practical assignment, which would contribute to each pupil's practical mark. Pupils were also informed that the recorded information would be analysed by them later; again in the form of a mathematics work sheet. (This was only done after all data were collected in order to prevent that the learning effect and increased food awareness affected the subsequent data collection). Worksheets on which learners performed curriculum-related data analyses (summation, rank ordering, calculation of means and proportions, and compilation of graphs) were set up for meaningful integration of nutrition and mathematics and for feedback. The intended learning outcomes and mathematics assignments based on the food records were jointly compiled by the researcher and the involved teacher.

3.2.1.2.1 Form

The format for the three-day food record resembled a simplified version of the one often used by the USDA in the Beltsville Human Nutrition Research Centre.[268] It was an open-ended form with each of the three days for recording printed on a different colour paper (Addendum B). For each recording day three A4-sized papers were printed with suitable headings. On the first page several examples were filled in. Each child also received written instructions.

3.2.1.2.2 Time frame and programme

In a two-week time span (10/9/2000 to 21/9/2000 that is early spring in South Africa), three recording periods of each three consecutive days (that is Sunday-Monday-Tuesday, Thursday-Friday-Saturday, and Tuesday-Wednesday-Thursday) were identified for documenting dietary intake. The programme schedule was set up in such a way that, for the sample as a whole, all days of the week were represented, and both, weekdays and weekend-days, formed the starting point for record keeping. This approach prevented recording fatigue and inaccurate reporting as observed by Gersovitz [50] to fall on a particular day. These researchers found a “significant association of actual and recorded values on the first two days of recording, but decreasing accuracy of recording afterwards”.

Thus, in total, intakes of seven weekdays and two weekend-days were measured. Tuesday and Thursday were each represented twice (Table 3.6).

TABLE 3.6: NUMBER OF TIMES THAT EACH DAY OF THE WEEK FORMED FIRST, SECOND OR THIRD RECORDING DAY

Day of week	**Day of recording**			
	First	Second	Third	Total
Sunday	1	0	0	1
Monday	0	1	0	1
Tuesday	1	0	1	2
Wednesday	0	1	0	1
Thursday	1	0	1	2
Friday	0	1	0	1
Saturday	0	0	1	1
Total	3	3	3	9

3.2.1.2.3 Sample composition and grouping

All 176 grade six pupils from an Afrikaans speaking, primarily middle-class, public primary school in urban Pretoria, South Africa, were included in the testing study. All pupils in the grade were white. Three classes were taught by the same mathematics teacher. From each of these, eleven pupils were randomly chosen from the alphabetic class list and allocated to the electronic scale group (only eleven electronic scales were available). These children were given a Soehnle digital scale to take home. From the remaining children those that had kitchen scales at home and who had parental permission to use these, were requested to record intakes using these scales. Children indicated the type of scale (spring, balance, digital) they had used on the record form. The rest of the children were provided a set of household measuring cups (250mL) and spoons (15mL tablespoons and 5mL teaspoons).

3.2.1.2.3 Training and briefing

The mathematics teacher who taught the three classes was trained to brief all the children. All data collection and briefing materials were provided to him, including:

(i) A <u>reference file</u> containing

- Copies of the data collection materials, that is
    - general and specific instructions
    - food record forms
    - food description flow diagrams
- Enlarged transparencies of the above, partially completed with relevant examples
- Class lists for recording groupings (according to type of quantification method used by each child [provided electronic scale, home scale or household measure] and for indicating which scale was issued to which child)
- Transparency pens

(ii) A <u>demonstration kit</u> for practically showing the correct use of the electronic scale containing:

- Electronic scale
- Plate and knife
- Slice of bread
- Margarine
- Cheese
- Coffee mug
- Apple

(All foods were in separate containers)

The training session focused on the following:

- Proper administrative and logistical matters
- Correct completion of the food record form, specifically in respect of describing food using the provided flow diagrams
- Appropriate use of the TARA function of the scale (practical demonstration: slice of bread first spread with margarine, then cheese; zeroing of scale after utensils, that is plates and mugs have been weighed.)

During the briefing six food description flow diagrams (meats, fats, bread, fluids, fruit / vegetables, snacks) similar to those proposed by Wold et al [156], but adjusted for South African circumstances were explained to the class using overhead transparencies. The diagrams were also included in each child's instruction folder for continuous referral in compiling the detailed description of foods eaten (see Addendum C). The children were requested to hand in wrappers of purchased foods in a plastic envelope, which was also part of the handout.

The children were instructed to bring their completed forms to school after the first day of recording to check for compliance.

3.2.1.2.5 Findings and conclusions

Each returned food record was rated by the researcher using comprehensiveness of completion of the "description of foods" column as only criterion. This eliminated discrimination based on type of quantification (that is type of scale or household measure, or provided versus own scales), potential measurement accuracy (again based on type of quantification method available to a child) or number of foods recorded (that is presumed completeness of record).

The marks given by the researcher are stated in the bottom row of Table 3.7.

Assessment of the analysis assignments was purely a school mathematics activity.

TABLE 3.7: NUMBER OF SUBJECTS IN THE THREE INTAKE QUANTIFICATION GROUPS, AS WELL AS MARKS ACHIEVED

Class and recording period		Quantification method			Total
		Electronic scale	Own non-electronic scale	Household measure	
Grade 6S (Sunday to Tuesday)		11	19	5	35 out of a class of 36
Grade 6T (Thursday to Saturday)		12	20	1	33 out of a class of 35
Grade 6P (Tuesday to Thursday)		10	18	4	32 of a class of 35
Grade 6E (Thursday to Saturday)		1	21	14	36 out of a class of 36
Grade 6H (Thursday to Saturday)		2	26	6	34 out of a class 35
Total		36	104	30	170 out of 176 learners
Marks for assignment (out of 10)	10	16 (44.4%)	11 (10.5%)	2 (6.7%)	29 (17.1%)
	9	9 (25.0%)	21 (20.0%)	2 (6.7%)	32 (18.8%)
	8	3 (8.0%)	22 (21.1%)	5 (1.7%)	30 (17.6%)
	=<7	8 (22.2%)	50 (48.0%)	21 (70.0%)	79 (46.5%)

Four children had electronic scales at home. In Table 3.7 these are indicated together with the handed out electronic scales. One of the randomly chosen children in this group fell ill, resulting in a total of 36 respondents who measured food intake electronically. Thirty children assessed their intake with household measures provided to them, and the rest (n=104) used home scales (all of which were reported to be spring type scales).

In total 61 learners returned detailed records (nine or more marks given for assignment). The food record of 30 pupils was rated eight out of ten in terms of comprehensiveness of food description, whilst 79 received seven marks or less. Six records were not returned (in two cases because of illness). This represents a response rate of 96.5%.

As evident from Table 3.7, the response rate was not related to a particular quantification method or class. However, the data quality (in terms of comprehensiveness of food descriptions) was clearly superior in the group that weighed their intake on electronic scales. Forty four percent of this group achieved full marks, whilst 70% of those who had been given household measures scored seven or less marks out of ten.

It was concluded that the electronic scales may have acted as a form of motivator and since the food record was supposed to be the reference method in the main study, it was decided to obtain more electronic scales (even though it was not financially possible to do this for all children). The finding from sub-study 3 strengthened this decision. Furthermore, for logistical reasons and for consistency in terms of administration it was decided to only use classes taught by one teacher in the main (validation) study.

3.2.2 Reference method 2: Screener by parents

For the second reference method, a text version of the dietary fat screener was compiled for completion by parents on behalf of their respondent child (Addendum D). This was very similar to the original NCEP tool (apart from the South African food examples within the food categories, the omission of group 2 sub-categories to match the test method, the translation to Afrikaans and layout to fit one page), and was therefore not subjected to developmental evaluation. Colleagues of the researcher checked understandability and technical aspects.

3.3 REFLECTION

As part of the confirmation that dietary assessment is never perfect, the developmental evaluation sub-studies on the test method greatly contributed to the understanding of some of the errors to be expected from children completing a FFQ type assessment. Where feasible, adjustments could be made and consequently the main study could be approached with confidence. This was strengthened by the knowledge that reference method 1, the three-day food record, was not only accepted, functional and standardised in the mathematics context, but that it could make a meaningful contribution to the curriculum.

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.

CHAPTER 5:RESULTS

Chapter overview

5.1 SAMPLE(S)

All 108 children on the class lists were anthropometrically assessed. Of these, 101 completed the screener and had no missing values in that data set. The re-test sample consisted of 39 children (three groups of each 13 learners). Four children did not participate in the food recording (parental or participant non-consent or absenteeism). Visual inspection of food records resulted in the following additional data cleaning: One food record was discarded because the child reported having gastro-enteritis during the recording period, five had omitted at least one full day of recording, one obviously under-recorded (wrappers were handed in but the foods were not recorded) and for one child one day had clearly been completed by a caregiver and the record appeared like a phantom report. This resulted in seven additional records being excluded from further analyses (some participants were excluded for more than one reason), leaving 93 children for whom the test method, complete anthropometry and usable food records (including three days of recording) were available.

A total of 78 parents returned a dietary screener completed in respect of their grade six child. Six of these contained at least one missing value and were not included in the comparative validation because the final scores would then reflect less than ten category scores. The eventual, triangulation-type comparison was based on this sample of 72 where complete information was available for the test method and both reference methods.

In Table 5.1 the composition of the various samples is summarised in terms of gender, age and anthropometric indices. Overall, the mean age was about twelve years and four months. Apart from height for age for boys, mean percentiles were higher than the median and mean z-scores were positive.

TABLE 5.1: DESCRIPTION OF THE VARIOUS SAMPLES IN TERMS OF MEAN ± STANDARD DEVIATION OF AGE AND ANTHROPOMETRIC INDICES

Sample		Age (Months)	Weight (kg)	Weight for age		Height (m)	Height for age		BMI (kg/m^2)	BMI for age	
				Percentile	Z-score		Percentile	Z-score		Percentile	Z-score
Test method	Male (n=50)	148.9±4.4	47.2±14.2	53.3±32.5	0.20±1.2	1.51±0.07	47.0±26.9	-0.09±0.8	20.4±4.9	58.6±32.0	0.36±1.2
	Female (n=58)	147.5±4.4	49.3±11.0	64.0±29.7	0.46±1.1	1.55±0.06	57.7±24.9	0.23±0.8	20.6±3.9	63.8±30.2	0.45±1.1
	Total (n=108)	148.1±4.4	48.3±12.6	59.0±31.3	0.34±1.1	1.53±0.06	52.6±26.3	0.08±0.8	20.5±4.4	61.4±31.0	0.41±1.1
Retest	Male (n=19)	148.5±4.4	50.3±16.0	60.7±35.0	0.46±1.4	1.51±0.08	50.0±29.9	-0.02±1.0	21.6±5.1	65.9±34.4	0.64±1.3
	Female (n=20)	147.8±3.5	47.4±9.7	59.6±28.7	0.30±0.9	1.54±0.05	55.0±25.2	0.16±0.7	19.8±3.5	58.7±29.8	0.25±1.0
	Total (n=39)	148.1±3.9	48.8±13.1	60.1±31.5	0.38±1.2	1.53±0.07	52.5±27.3	0.07±0.9	20.7±4.4	62.2±31.9	0.38±1.2
Comparative validation 1 (Food record as reference method)	Male (n=43)	149.0±4.6	48.0±16.4	55.3±32.4	0.26±1.3	1.51±0.07	48.6±26.9	-0.04±0.8	20.6±5.1	59.5±32.7	0.40±1.2
	Female (n=50)	147.5±3.9	49.7±11.1	64.8±29.1	0.49±1.1	1.55±0.06	58.4±25.1	0.26±0.8	20.6±3.9	63.5±30.2	0.43±1.1
	Total (n=93)	148.2±4.3	48.9±12.8	60.4±30.9	0.38±1.1	1.53±0.07	53.9±26.3	0.12±0.8	20.6±4.5	61.7±31.1	0.42±1.1
Comparative validation 2 (Parents as reference method)	Male (n=36)	148.7±4.3	49.1±15.3	57.9±33.0	0.36±1.3	1.52±0.08	50.1±29.2	-0.01±0.9	21.1±5.3	62.3±32.7	0.51±1.2
	Female (n=42)	147.7±3.5	48.8±10.2	63.3±28.3	0.43±0.9	1.55±0.06	56.5±25.2	0.19±0.8	20.3±3.4	62.7±28.1	0.42±0.9
	Total (n=78)	148.2±3.9	48.9±12.7	60.8±30.4	0.40±1.1	1.53±0.07	53.5±27.1	0.09±0.9	20.7±4.4	62.5±30.1	0.47±1.1
Triangulation	Male (n=34)	148.0±4.7	46.9±10.7	57.7±31.9	0.24±1.1	1.53±0.06	53.9±26.7	0.11±0.82	19.9±3.8	57.5±32.0	0.26±1.1
	Female (n=38)	149.3±4.1	52.7±14.5	67.1±29.3	0.67±1.2	1.54±0.06	54.6±26.0	0.14±0.80	22.0±5.1	70.6±29.6	0.73±1.2
	Total (n=72)	148.6±4.4	50.0±13.1	62.7±30.7	0.47±1.2	1.54±0.06	54.3±26.2	0.12±0.8	21.0±4.6	64.4±31.2	0.51±1.2

5.2 TEST METHOD

5.2.1 Internal consistency

The item total correlations (Table 5.2) ranged from 0.35 for table fats to 0.66 for cheese. All were highly significant (P<0.0001 in most cases). Cronbach's coefficient alpha for all ten category scores in the whole sample was 0.70 for the raw variables. When performed with deleted variables (that is without table fats), alpha increased to 0.72. The split half method of estimating reliability yielded a correlation coefficient between the two parts of the screener of 0.57 (P<0.0001).

TABLE 5.2: CORRELATION COEFFICIENTS (r) BETWEEN THE CATEGORY SCORES OF ALL FOOD CATEGORIES AND THE FINAL SCORES (n=101)

Food category	r[a]
Meat	0.55[b]
Eggs	0.41[b]
Dairy, milk, high fat	0.54[b]
Dairy, cheese, high fat	0.66[b]
Dairy, dessert, high fat	0.55[b]
Fried foods	0.38[b]
In baked goods	0.64[b]
Convenience foods	0.51[b]
Table fats, high fat	0.35[c]
Snacks, high fat	0.65[b]

[a] Pearson's correlation coefficients
[b] P<0.0001
[c] P<0.0004

5.2.2 Test-retest reproducibility

The check for sampling bias revealed that, within the first administration, there was no significant difference between repeaters and non-repeaters (P>0.05 for all the category scores as well as the final scores; Table 5.3). Cronbach's alpha for the re-test sample was 0.67 for all raw variables and 0.69 with meat deleted.

TABLE 5.3: DIFFERENCE IN CATEGORY AND FINAL SCORES BETWEEN PARTICIPANTS IN RE-TEST AND NON-PARTICIPANTS

Food category	Category scores Mean±SD		P[a]
	Participants (n=39)	Non-participants (n=65)	
Meat	13.2±5.9	13.2±6.3	0.95
Eggs	4.8±3.1	4.8±4.0	0.87
Dairy, milk, high fat	14.4±5.4	14.1±6.7	0.94
Dairy, cheese, high fat	11.7±5.9	9.3±6.2	0.05
Dairy, dessert, high fat	10.6±6.0	10.5±6.5	0.91
Fried foods	11.8±6.6	12.0±6.8	0.93
In baked foods	8.7±4.9	8.58±6.2	0.88
Convenience foods	10.6±6.1	9.5±6.2	0.35
Table fats, high fat	15.8±6.2	14.8±7.1	0.58
Snacks, high fat	12.5±5.7	11.7±6.6	0.52
Final score	114.1±24.9	108.1±35.1	0.42

[a]Wilcoxon Two-sided Rank Sum test

Table 5.4 shows the degree of agreement in the two administrations in terms of the portion size estimates and the categorised frequencies of intake. The percentage of children reporting the identical usual portion size in both administrations varied from less than 50% for milk and snacks to over 70% for dessert, eggs, baked goods, convenience foods, table fats and fried foods. Following adjustment for chance agreement (kappa statistic), table fats, convenience foods and eggs had moderate agreement (kappa 0.41-0.60). Meat, baked goods and dessert showed fair agreement (kappa 0.21-0.40) and for the remaining four food categories agreement was poor (kappa <0.20).[270] The McNemar statistic revealed a departure of symmetry (P<0.05) only for meat, suggesting that, in general, changes in reported portion sizes were similar in both directions (from smaller to larger and vice versa) in the second administration.

TABLE 5.4 REPRODUCIBILITY OF PORTION SIZE AND FREQUENCY OF INTAKE ESTIMATES OF ALL FOOD CATEGORIES (n=39)

Food categories	Portion size				Frequency of intake			
	Identical (%)	Kappa		McNemar P	Identical (%)	Kappa		McNemar P
		Value	P			Value	P	
Meat	65.8	0.36	0.01	0.04	82.1	0.53	0.0001	0.06
Eggs	70.6	0.43	0.01	0.37	82.1	0.55	0.0001	0.51
Dairy, milk, high fat	43.2	0.10	0.40	0.44	87.2	0.12	0.30	0.51
Dairy, cheese, high fat	65.7	-0.15	0.26	0.84	48.7	0.14	0.21	0.06
Dairy, dessert, high fat	70.3	0.28	0.16	0.37	56.4	0.19	0.11	0.35
Fried foods	76.3	0.19	0.19	0.10	48.7	0.08	0.53	0.72
In baked goods	71.8	0.34	0.004	0.15	61.5	0.36	0.002	0.04
Convenience foods	72.2	0.48	0.001	0.26	61.6	0.30	0.02	0.51
Table fats, high fat	75.0	0.52	0.002	0.32	84.6	0.17	0.27	0.41
Snacks, high fat	48.7	0.08	0.51	0.62	76.9	0.46	0.002	0.34

In the case of estimated weekly intake, the percentage of agreement ranged from just under 50% for cheese and fried foods to over 80% for meat, eggs, table fats and milk. Correction for chance agreement resulted in moderate agreement for eggs, meat and snacks (kappa 0.41-0.60), fair agreement for baked goods and convenience foods (kappa 0.21-0.40) and poor agreement for the remaining five food categories. Baked goods, cheese and meat were borderline (P about 0.05) in terms of symmetry of the non-identical responses in the two administrations, but, in general, for reported weekly consumption, the increases seemed to be balanced by the decreases (see McNemar information in Table 5.4).

As evident from Table 5.5, the mean category scores in the second administration were lower than in the first administration for six of the ten food categories, but the difference was non-significant (P>0.05). Meat and cheese were clear exceptions, with fats in baked foods being borderline. The mean final scores in the two administrations also did not differ significantly. From the frequency distribution of the difference between the final scores in the two administrations, it was found that for the whole group about 72% of children were within plus or minus 30 points (Table 5.6). For boys this corresponding cumulative percentage was 58, whilst for girls it was 85. The percentage of girls and boys was similar for an absolute difference in final scores of 20 or less. There was no significant correlation between BMI for age Z-scores and difference in final scores in the two administrations, neither for the group as a whole (r=0.05, P=0.75), nor for the genders separately (males r=-0.08, P=0.74; females r=0.33, P=0.16).

TABLE 5.5 MEAN ±SD CATEGORY AND FINAL SCORES IN THE FIRST AND SECOND ADMINISTRATION (n=39)

Food category		Category scores Mean ± standard deviation			P [c]
		First administration	Second administration	Difference [a]	
Meat		13.2±5.9	16.2±5.9	-2.9±6.0	0.01
Eggs		4.8±3.1	4.4±3.6	0.4±2.9	0.47
Dairy, milk, high fat		14.4±5.4	14.0±6.2	0.4±7.3	0.79
Dairy, cheese, high fat		11.7±5.9	8.6±5.9	3.1±7.1	0.01
Dairy, dessert, high fat		10.6±6.0	10.2±6.5	0.4±7.7	0.76
Fried foods		11.8±6.6	12.8±6.8	-1.0±9.2	0.57
In baked foods		8.7±4.9	10.4±7.2	-1.7±6.2	0.05
Convenience foods		10.6±6.1	9.4±6.1	1.2±8.0	0.48
Table fats, regular (high fat)		15.8±6.2	14.4±6.5	1.3±7.0	0.35
Snacks, high fat		12.5±5.7	12.9±6.6	-0.5±6.7	0.52
Final score	Whole group (CI[b])	114.1±24.9 (106.0, 122.2)	113.4±31.1 (103.3, 123.5)	0.69±32.6 (-9.9, 11.3)	0.86
	Males (CI[b])	115.5±30.6 (100.7, 130.2)	117.2±31.6 (101.9, 132.4)	-1.68±39.4 (-20.7, 17.3)	0.82
	Females (CI[b])	112.8±18.7 (104.0, 121.5)	109.8±31.0 (95.3, 124.3)	3.0±25.4 (-8.9, 14.8)	0.78

[a] First minus second administration
[b] 95% Confidence limits containing the mean
[c] Wilcoxon's Signed Rank test

TABLE 5.6: ABSOLUTE FINAL SCORE DIFFERENCE BETWEEN ADMINISTRATIONS, BY GENDER (n=39)

Absolute final score difference	Respondents								
	Whole group			Males			Females		
	n	%	Cumulative %	n	%	Cumulative %	n	%	Cumulative %
≤10	11	28.2	28.2	5	26.3	26.3	6	30.0	30.0
>10 but ≤20	6	15.4	43.6	4	21.1	47.4	2	10.0	40.0
>20 but ≤30	11	28.2	71.8	2	10.5	57.9	9	45.0	85.0
>30	11	28.2	100.0	8	42.1	100.0	3	15.0	100.0
Total	39	100.0		19	100.0		20	100.0	

An initial indication of a linear relationship between the final scores of the two administrations was suggested by small, but statistically significant correlation coefficient ($r=0.36$, $P=0.02$). If separated by gender, the correlation coefficient for boys was non-significant ($r=0.26$, $P=0.29$) whilst it was highly significant for girls ($r=0.58$, $P=0.01$). Measures of central tendency and variability in the final scores of the dietary fat screener are presented in Table 5.7. For the whole group the means were very similar, but the standard deviations, inter-quartile ranges and 95% confidence intervals point to variability. Girls as a group exhibited overall less variability and the midpoints of the second administration were lower than the first, in contrast to the boys. From Table 5.5 (bottom row) it is, however, evident that for boys and girls the difference in final scores did not significantly differ from zero.

TABLE 5.7: MEASURES OF LOCATION AND VARIABILITY OF FINAL SCORES OF THE DIETARY FAT SCREENER (n=39; 19 male, 20 female)

Measure		First administration	Second administration	Difference[a]
Mean ± standard deviation	Whole group	114.1±24.9	113.4±31.1	0.69±32.6
	Males	115.5±30.6	117.2±31.6	-1.68±39.4
	Females	112.8±18.7	109.8±31.0	2.95±25.4
Median (P_{25}, P_{75})[b]	Whole group	112.0 (95, 133)	104.0 (93, 142)	1.0 (-20, 26)
	Males	105.0 (87,146)	105.00 (100, 142)	-1.00 (-20, 29)
	Females	112.0 (99.0, 125.5)	100.5 (90.5, 137.0)	5.00 (-22, 24)
95% Confidence interval	Whole group	(106.0, 122.2)	(103.3, 123.5)	(-9.9, 11.3)
	Males	(100.7, 130.2)	(101.9, 132.3)	(-20.7, 17.3)
	Females	(104.0, 121.5)	(95.3, 124.3)	(-8.9, 14.8)

[a] First minus second administration
[b] P_{25} = 25th percentile, P_{75} = 75th percentile

From Table 5.8 it can be deducted that 40% (3 plus 5) of children remained in the same lowest or highest quarter during the two administrations, whilst 15% (2 plus1) changed from one extreme to the other. When the scores were categorised, then less than 8% (n=3; 2 female, 1 male) of respondents were reclassified from a high fat to a prudent diet.

TABLE 5.8: NUMBER OF RESPONDENTS IN OPPOSITE QUARTERS IN FIRST AND SECOND ADMINISTRATION (n per quartile = 10)

		First administration	
		Lowest quarter	Highest quarter
Second administration	Lowest quarter	3	1
	Highest quarter	2	5

Figures 5.1a and 5.1b quantify and illustrate the variability of the individual data points in the two administrations. Visual inspection confirms a wide scatter around the diagonal line of equality (Figure 5.1a) and the horizontal line of zero difference (Figure 5.1b) for the group as a whole, with females tending to be closer to these lines, suggesting better reproducibility.

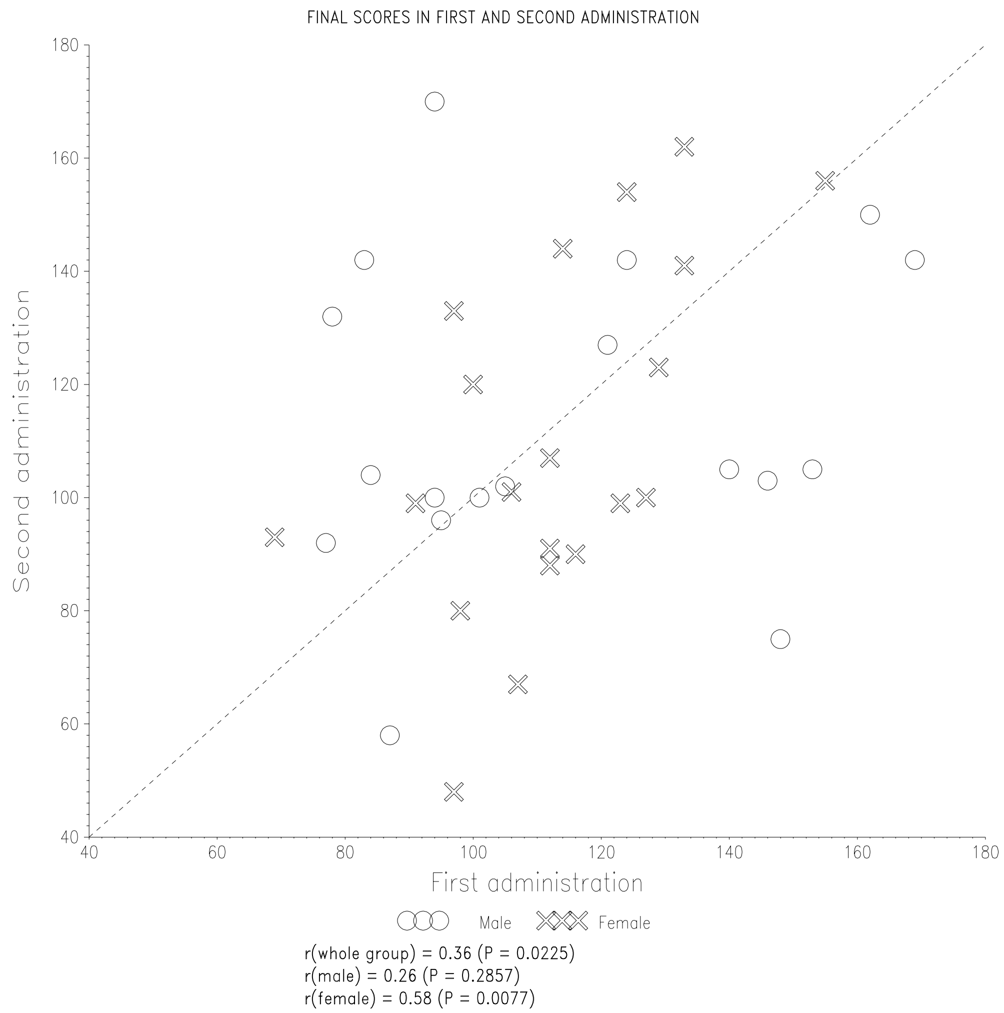


FIGURE 5.1a: FINAL SCORES IN FIRST AND SECOND ADMINISTRATION

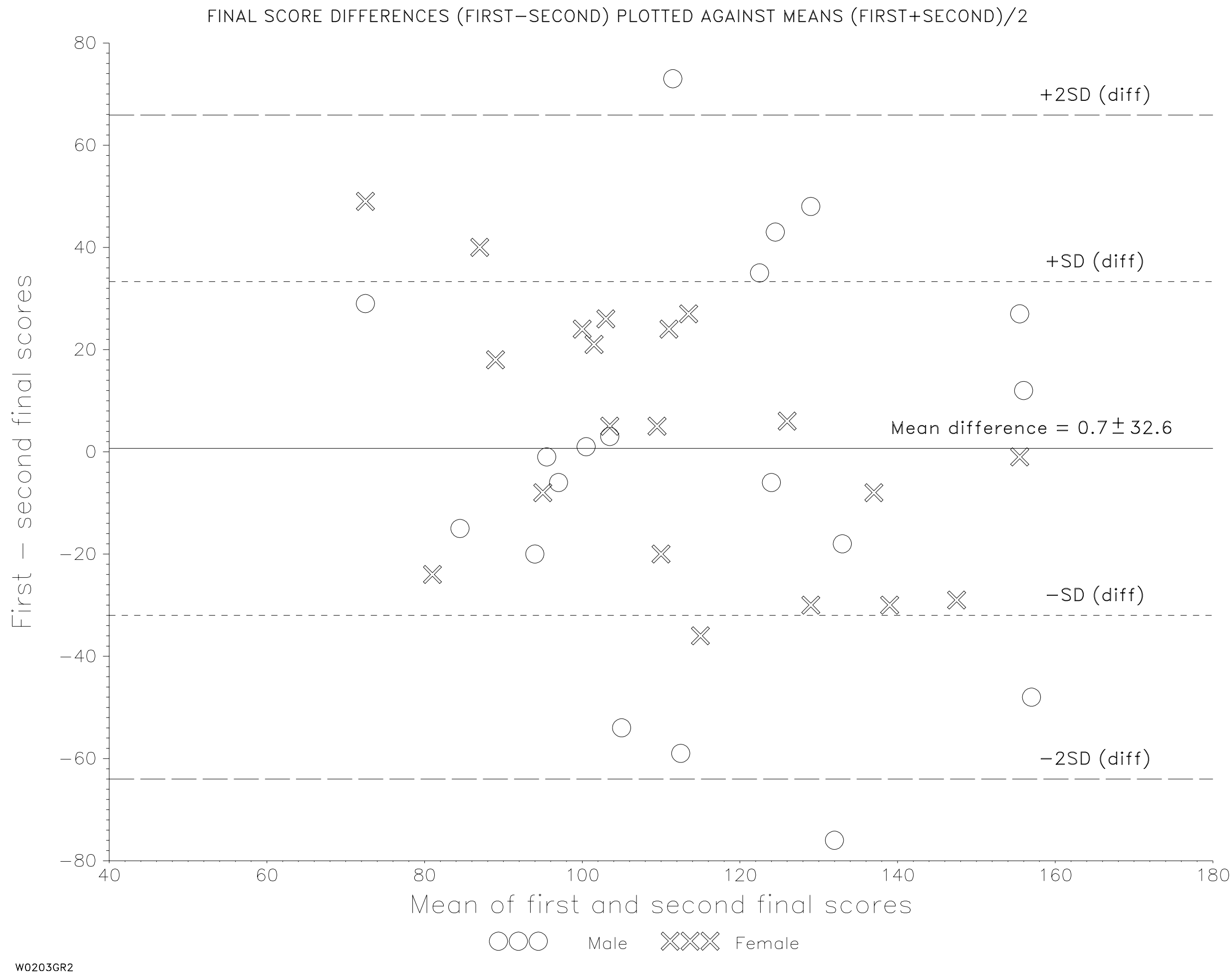


FIGURE 5.1b: **FINAL SCORE DIFFERENCES (FIRST – SECOND ADMINISTRATION) PLOTTED AGAINST MEANS ([FIRST + SECOND]/2)**

5.3 FOOD RECORD

5.3.1 Quality control

The mean PAL for the final, whole group (n=93) was 1.45±0.4 (range 0.7-2.8). For boys and girls separately the respective values were 1.49±0.4 and 1.41±0.4. Twelve children had a PAL value below 1.06. On average the recorded mean daily energy intakes were 97% and 91% of the 2002 DRI values [10] for boys and girls respectively.

For the group as a whole, a statistically significant positive correlation (r=0.24; P=0.02) was found between mean energy intake over the three days and body weight. For boys this correlation coefficient was 0.46 (P=0.002) and for girls it was shown to be negative (r=-0.12) though not statistically significant (P=0.42).

The energy intakes over the three days, individually and as a group, for each of the three recording periods as well as for the two quantification methods (electronic scale and household measures) are indicated in Table 5.9. Overall there was a statistically significant difference in energy intake between the three recording days (P=0.03) in the sense that the z statistic was larger than the critical value of alpha of 0.05 when day one was compared to day three. Within the Thursday to Saturday and the Sunday to Tuesday recording periods there was no significant difference in energy intake over the three days, but in the Tuesday to Thursday group the difference was statistically significant (P=0.04). Again, specifically between day one and day three the z statistic exceeded the critical value for overall alpha of 0.05.

TABLE 5.9: MEAN (±SD) ENERGY INTAKES OVER THE THREE RECORDING DAYS BY RECORDING PERIOD AND QUANTIFICATION METHOD (n=93)

		Energy intake (kJ)			
		Day 1	**Day 2**	**Day 3**	**Means**
Recording period	Thursday - Saturday (n=22)	8999±2539	8585±2422	9900±7357	9161±7357
	Tuesday - Thursday (n=49)	9562±3735	8460±2818	8204±2897	8742±2505
	Sunday - Tuesday (n=22)	8201±1888	7201±2460	6989±2154	7464±1506
Quantification method	Electronic scales (n=60)	8840±2985	7946±2504	7776±3643	8187±2270
	Household measures (n=33)	9592±3401	8640±2955	9304±5300	9179±2794
	Means	9107±3142	8192±2678	8318±4338	8539±2499

5.3.2 Energy and fat intakes

Reference method one, the three-day food record, contained three measures of high fat intake: Mean percentage total daily fat energy (PFE), mean percentage daily saturated fat energy (PSFE) and mean daily cholesterol intake. In Table 5.10 the mean dietary intakes of these three measures as well as energy intake over the three days of recording are given. In addition the International Institute of Medicine’s estimates of within-subject variation for the corresponding nutrients is stated for comparing the observed day-to-day variability in intake to international ‘standards’.

TABLE 5.10: MEAN (±SD[a]) INTAKES (THREE-DAY RECORD) AND PUBLISHED INTRA-SUBJECT VARIATION FOR 9-18 YEAR OLDS [272]

Intake	Boys			Girls			Whole group (n=93)
	This study (n=43)	Intra-subject variation		This study (n=50)	Intra-subject variation		
		SD	CV[b] (%)		SD	CV (%)	
Energy (kJ)	9280±3032	3360	33	7902±1717	2638	34	8539±2499
Total fat (g)	86.9±36.0	38.2	42	74.3±21.2	29.8	45	80.1±29.6
PFE[c]	34.9±6.2	NA	NA	35.5±6.5	NA	NA	35.2±6.3
Saturated fatty acids (g)	32.0±15.1	15.3	48	26.7±8.3	11.3	48	29.1±12.2
PSFE[d]	12.9±2.9	NA	NA	12.8±3.1	NA	NA	12.8±3.0
Cholesterol (mg)	275.7±128.0	199	71	211.6±99.2	145	72	241.3±117.3

[a] Standard deviation
[b] Coefficient of variation
[c] Percent fat energy
[c] Percent saturated fat energy
NA: Not applicable

Seventy-two (77.4%) of the children consumed a diet where PFE contributed more than 30% of the energy intake. Seventy-eight (83.9%) and 19 (20.4%) respectively recorded a PSFE higher than ten percent, and mean daily cholesterol consumption greater or equal to 300mg. Figure 5.2 shows that there was considerable overlap between the measures of high fat intake. Sixteen (17.2%) participants were classified as high fat consumers by *all* three measures, whilst 84 (90.3%) of participants would be classified as having a high fat intake if *any* one of the measures was used as criterion. Nine participants' intake was classified as prudent by all three measures.

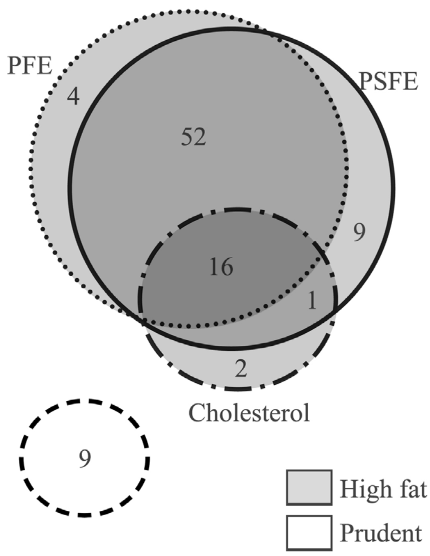


FIGURE 5.2: CLASSIFICATION OF FOOD RECORDS INTO 'HIGH FAT' AND 'PRUDENT' USING PERCENT FAT ENERGY (PFE), PERCENT SATURATED FATTY ACID ENERGY (PSFE) OR CHOLESTEROL INTAKE AS MEASURES (n=93)

5.3.3 Comparative validation: Test method versus food record

The correlations between the final score obtained in the screener and the three measures of fat intake from the three-day food record are presented in Table 5.11. The Table shows that for girls, the final score of the screener was significantly ($P<0.05$) related to total fat energy, total saturated fat energy and cholesterol intake. For the group as a whole and for boys no one of the measures was significantly correlated. For girls the associations with PFE and PSFE were also non-significant.

TABLE 5.11: CORRELATION COEFFICIENTS (r) BETWEEN TEST METHOD FINAL SCORES AND FOOD RECORD MEASURES OF FAT INTAKE

Measure of fat intake	**Boys** (n=43)		**Girls** (n=50)		**Whole group** (n=93)	
	r	P	r	P	R	P
Total fat energy	0.03	0.87	0.30	0.04	0.16	0.13
PFE [a]	-0.17	0.28	0.14	0.35	-0.02	0.82
Saturated fat energy	0.06	0.72	0.31	0.03	0.20	0.06
PSFE [b]	-0.01	0.93	0.19	0.18	0.08	0.45
Cholesterol	-0.09	0.56	0.31	0.03	0.18	0.10

[a] Percentage fat energy
[b] Percentage saturated fatty acid energy

The mean energy and fat intakes of those classified as consuming a high fat versus a prudent diet according to the test method are presented in Table 5.12.

TABLE 5.12: MEAN (±SD) INTAKES IN TEST METHOD CLASSIFCATION GROUPS (n=93)

Food record measure	**Test method classification**	
	High fat (n=86)	Prudent (n=7)
Energy (kJ)	8589±2521	7928±2303
Total fat (g)	80.5±29.5	75.7±32.3
PFE [a]	35.2±6.4	35.2±5.2
Saturated fat (g)	29.3±12.0	26.8±14.8
PSFE [b]	12.9±3.0	12.2±3.5
Cholesterol (mg)	246.0±113.1	183.1±158.8

[a] Percentage fat energy
[b] Percentage saturated fatty acid energy

Table 5.13 is a multiple cross tabulation of classifications based on each of the measures of fat intake from the food record on the one hand, with the classification according to the screener, on the other hand.

TABLE 5.13: COMPARISON OF TEST METHOD CLASSIFICATION TO CLASSIFICATION OF THREE MEASURES FROM FOOD RECORD (n=93)

		Three-day food record measures									
		PFE[a]		PSFE[b]		CHOL[c]		ANY[d]		ALL[e]	
		+ n(%)	- n(%)	+ n(%)	- n(%)	+ n(%)	- n(%)	+ n(%)	- n(%)	+ n(%)	- n(%)
Test method classification[f]	+ n=86	66 (71.0)	20 (21.5)	74 (79.6)	12 (12.9)	18 (19.4)	68 (73.1)	78 (83.9)	8 (8.6)	15 (16.1)	71 (76.3)
	- n=7	6 (6.6)	1 (1.1)	4 (4.3)	3 (3.2)	1 (1.1)	6 (6.5)	6 (6.5)	1 (1.1)	1 (1.1)	6 (6.6)
Total n (%)	93 (100)	72 (77.4)	21 (22.6)	78 (83.9)	15 (16.1)	19 (20.4)	74 (79.6)	84 (90.3)	9 (9.7)	16 (17.2)	77 (82.8)

a Percentage fat energy, where " + " is >30% (high fat); " - " is =<30% (prudent)
b Percentage saturated fatty acid energy, where " + " is >10% (high fat); " - " is =<10% (prudent)
c Cholesterol intake, where " + " is >=300mg (high fat); " - " is <300mg (prudent)
d Any one of the three measures applies
e All three measures apply simultaneously
f " + " is final score >68 (high fat); " - " is final score =<68 (prudent)

When PFE, PSFE and any measure acted as reference, the percentage true positives (high fat intake) plus true negatives (prudent fat intake) was high. Thus the percentage exact matches (overall predictive value) for PFE, PSFE and "ANY" were 72%, 83% and 85% respectively. This was in contrast to cholesterol intake or meeting all three measures as criterion, where many false positives were found. Nevertheless, the simple kappa coefficient was always below 0.20 denoting poor chance corrected agreement.

With the information from Table 5.13 as starting point, various indicators of the comparative validity of the screener relative to the three measures of fat intake are shown in Table 5.14.

TABLE 5.14: INDICATORS OF VALIDITY OF TEST METHOD AGAINST FOOD RECORD MEASURES OF FAT INTAKE (n=93)

Food record measure	Indicators of validity								
	Sens[g] (CI)[f]	Spec[h] (CI)[f]	OPV[i]	PPV[j] (CI)[f]	NPV[k] (CI)[f]	RR[l]	OR[m]	LR+[n] (CI)[f]	LR-[o] (CI)[f]
PFE[a]	0.92 (0.83, 0.96)	0.05 (0.00, 0.23)	0.72	0.77 (0.69, 0.84)	0.14 (0.01, 0.51)	5.37	0.55	0.96 (0.86, 1.08)	1.75 (0.22, 13.74)
PSFE[b]	0.95 (0.88, 0.98)	0.20 (0.07, 0.45)	0.83	0.86 (0.77, 0.92)	0.43 (0.16, 0.75)	2.01	4.63	1.20 (0.92, 1.54)	0.27 (0.06, 1.03)
CHOL[c]	0.95 (0.75, 1.00)	0.08 (0.04, 0.17)	0.26	0.21 (0.14, 0.31)	0.86 (0.49, 0.99)	0.24	1.59	1.03 (0.91, 1.17)	0.65 (0.83, 5.07)
ANY[d]	0.93 (0.85, 0.98)	0.11 (0.01, 0.44)	0.85	0.91 (0.83, 0.95)	0.14 (0.01, 0.51)	6.35	1.63	1.05 (0.82, 1.33)	0.64 (0.09, 4.76)
ALL[e]	0.94 (0.72, 1.00)	0.08 (0.04, 0.16)	0.23	0.17 (0.11, 0.27)	0.86 (0.49, 0.99)	0.20	1.27	1.02 (0.88, 1.17)	0.80 (0.10, 6.21)

a Percentage fat energy
b Percentage saturated fatty acid energy
c Cholesterol intake
d Any one of the measures of fat intake
e All measures of fat intake
f 95% confidence interval
g Sensitivity
h Specificity
i Overall predictive value
j Positive predictive value
k Negative predictive value
l Relative risk
m Odds ratio
n Positive likelihood ratio
o Negative likelihood ratio

The sensitivity of all three measures, individually or combined, was always very high (>0.9). The specificity was always low, with PSFE having the highest specificity, namely 0.20. The overall and the positive predictive values were higher than 0.7 for PFE, PSFE and when any of the three measures met the cut-off, but for cholesterol and when all three measures had to be met, they were less than 0.3. Negative predictive value was highest (0.86) for both, cholesterol and when all three measures applied. The relative risk was strong (that is above two)[45] for PFE, PSFE and when any one of the measures was applied. The odds ratio for PSFE was much higher than any one of the other two measures or combinations.

5.4 SCREENER BY PARENTS

5.4.1 Internal consistency

When parents completed the screener to assess their grade six children's diets the following was found: Apart from eggs, item total correlations were highly significant (P<0.0001) for all food categories (Table 5.15). For the latter nine food categories the correlation coefficients ranged from 0.48 (meat) to 0.63 (milk). Cronbach's coefficient alpha was 0.69 for all ten category scores and 0.71 when performed without the category scores for eggs. The correlation coefficient (Pearson) obtained between category scores of two random halves of the screener was 0.48 (P<0.0001).

TABLE 5.15 CORRELATION COEFFICIENTS (r) BETWEEN PARENTS' CATEGORY SCORES OF ALL FOOD CATEGORIES AND THEIR FINAL SCORES (n=76)

Food categories	R
Meat	0.48[a]
Eggs	0.09[b]
Dairy, milk, high fat	0.63[a]
Dairy, cheese, high fat	0.50[a]
Dairy, dessert, high fat	0.64[a]
Fried foods	0.52[a]
In baked goods	0.61[a]
Convenience foods	0.54[a]
Table fats, high fat	0.50[a]
Snacks, high fat	0.59[a]

[a] P<0.0001
[b] P=0.4650

5.4.2 Comparative validation: Test method versus screener by parents

5.4.2.1 Portion size

The percentage exact agreements between parents and children in terms of reported usual portion size eaten by the children varied from as low as 18% for fried foods to about 72% for eggs (Table 5.16) with an average percentage agreement over the ten food categories of 45.0%.

Chance corrected agreement based on kappa values was moderate for eggs (kappa=0.44; P=0.0001) and fair for cheese (kappa=0.30; P=0.0006). For all other food categories agreement was poor (kappa < 0.20). Apart from eggs and baked foods, the non-agreeing responses tended to be non-symmetrical (McNemar P<0.05).

5.4.2.2 Frequency of intake

From Table 5.16 it is also evident that the percentage child-parent pairs that reported identical frequency of intake ranged from 42% (convenience foods) to 74% (milk). This represents an average of 60.1% across all ten food categories. For 40% of the food categories (eggs, dessert, cheese and fried foods) the chance corrected agreement was fair (kappa between 0.21 and 0.40 and P<0.05). For the remaining six food categories the chance corrected agreement was poor. For 80% of food categories non-agreement around the diagonal of perfect matches appeared to be not symmetrical (P<0.05) with eggs and snacks being the exception, where P=0.91 and 0.25 respectively.

TABLE 5.16: AGREEMENT BETWEEN PARENTS AND CHILDREN IN TERMS OF REPORTED PORTION SIZE AND FREQUENCY OF INTAKE ESTIMATES OF ALL FOOD CATEGORIES (n=77)

Food categories	Portion size				Frequency of intake			
	Identical (%)	Kappa		McNemar P	Identical (%)	Kappa		McNemar P
		Value	P			Value	P	
Meat	46.7	0.07	0.42	0.01	68.8	0.07	0.36	0.00
Eggs	71.6	0.44	0.00	0.18	68.8	0.22	0.01	0.91
Dairy, milk, high fat	49.3	0.17	0.05	0.07	74.0	0.21	0.01	0.04
Dairy, cheese, high fat	63.2	0.30	0.00	0.05	59.7	0.32	0.00	0.00
Dairy, dessert, high fat	31.7	0.05	0.37	0.00	56.9	0.00	0.98	0.00
Fried foods	18.4	-0.08	0.07	0.00	58.4	0.21	0.03	0.05
In baked goods	53.7	0.14	0.12	0.20	48.1	0.05	0.56	0.10
Convenience foods	39.7	0.03	0.60	0.00	41.6	-0.06	0.38	0.00
Table fats, high fat	29.2	-0.02	0.73	0.00	66.2	0.05	0.34	0.00
Snacks, high fat	46.7	0.14	0.04	0.00	58.4	0.18	0.10	0.25

5.4.2.3 Category and final scores

Table 5.17 summarises the mean category scores of parents and children, as well as the correlations between the scores. Mean category scores of parents were always lower than those of the children, and in the case of dessert and convenience foods the mean was about half of that for the children. A statistically significant (P<0.05) linear relationship between parents' and children's category scores was found for meat, milk and cheese. In the case of table fats and snacks this relationship bordered on statistical significance (P more or less 0.05) whilst for eggs, dessert, fried, baked and convenience foods there was an absence of evidence for a linear relationship.

TABLE 5.17: CHILDREN'S AND PARENTS' CATEGORY AND FINAL SCORES [a]: MEANS AND CORRELATIONS (n=78)

Food category		Mean ±SD	Correlations	
			r	P
Meat	Children	13.4±6.2	0.25	0.03
	Parents	13.2±4.9		
Eggs	Children	5.2±3.9	0.12	0.28
	Parents	4.4±3.1		
Dairy, milk, high fat	Children	13.9±6.2	0.40	0.00
	Parents	11.2±6.4		
Dairy, cheese, high fat	Children	10.4±6.4	0.41	0.00
	Parents	8.2±5.5		
Dairy, dessert, high fat	Children	10.8±5.2	-0.03	0.84
	Parents	5.3±3.8		
Fried foods	Children	12.6±6.9	0.15	0.20
	Parents	8.4±4.6		
In baked goods	Children	8.6±5.2	0.09	0.42
	Parents	7.0±5.2		
Convenience foods	Children	10.8±6.4	0.04	0.70
	Parents	5.0±3.3		
Table fats, high fat	Children	16.1±6.2	0.22	0.05
	Parents	9.9±6.2		
Snacks, high fat	Children	12.3±6.1	0.22	0.06
	Parents	9.1±5.0		
Final score	Children	114.4±30.7	0.23	0.04
	Parents	82.0±25.2		

[a] Frequency score multiplied by portion size score.
Frequency score based on weekly consumption:
Rarely or never = 0
Up to 3 times = 3
More than 3 =7
Portion size score:
Small = 1
Medium = 2
Large = 3

From Table 5.17 it can be seen that the correlation coefficient between the final scores of children and their parents was 0.23 (P=0.04). For boys the correlation coefficient was non-significant (r=0.13; P=0.46) whilst for girls it higher than the whole group value (r=0.33; P=0.04).

Table 5.18 illustrates the mean difference between children's and parents' final scores. Whilst the mean difference between children and parents was very similar for boys and girls, greater variability (see standard deviations and confidence intervals) was apparent in the case of boys. Nevertheless, the difference in final scores between children and parents differed significantly from zero for the group as a whole and for the gender separately (P<0.0001 in all three instances).

TABLE 5.18: MEAN DIFFERENCE IN FINAL SCORES OF CHILDREN AND PARENTS BY GENDER

	Difference[a] Mean±SD	CI[b]	P[c]
Whole group	32.0±35.1	(23.8; 40.2)	0.0001
Boys	31.6±42.4	(17.0; 46.2)	0.0001
Girls	32.4±27.9	(23.3; 42.4)	0.0001

[a] Child minus parent
[b] 95% Confidence limits containing the mean difference
[c] Wilcoxon's Signed Rank Test

Figures 5.3a and 5.3b compare final scores obtained by children and by their parents. The plot in Figure 5.3a illustrates that in most cases the children had a higher final score than their parents as most of the data points were above the diagonal of perfect agreement. Furthermore this scatter plot gives a visual indication of why the correlation coefficient revealed the weak linear association reported above, in fact the absence of such an association for boys. Figure 5.3b re-emphasises the variability (random error) and the presence of bias (systematic error) by the magnitude of the standard deviation and mean difference respectively.

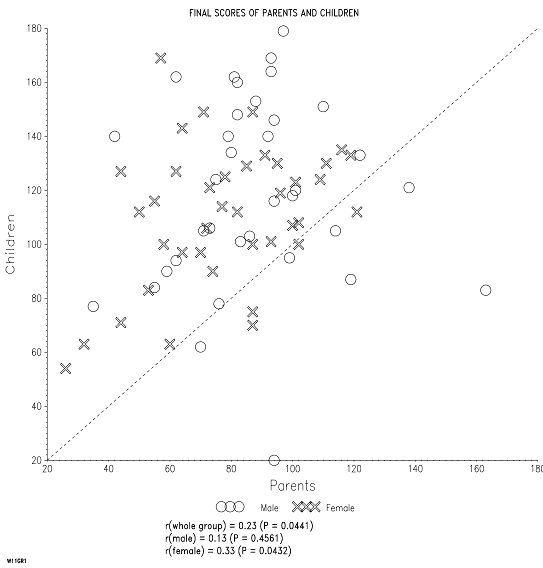


FIGURE 5.3a: FINAL SCORES OF PARENTS AND CHILDREN

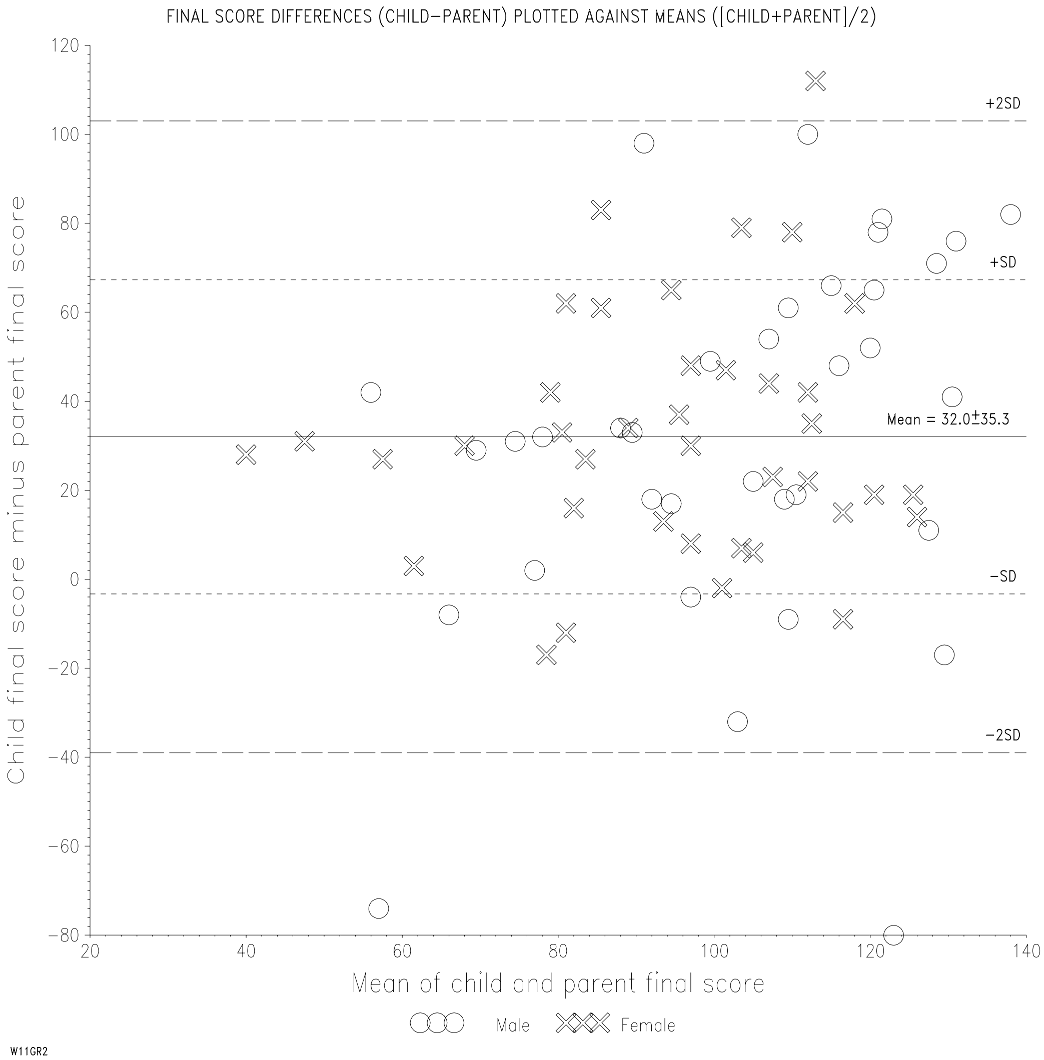


FIGURE 5.3b: FINAL SCORE DIFFERENCES (CHILD-PARENT) PLOTTED AGAINST MEANS ([CHILD + PARENT]/2)

5.4.2.4 Classification agreement

The percentage identical classifications into high fat or prudent was 76%. When corrected for chance the agreement was, however, poor (kappa = 0.16) and the McNemar test indicated a departure from symmetry (P=0.0010).

5.5 TRIANGULATION: TEST METHOD VERSUS FOOD RECORD VERSUS SCREENER BY PARENTS

Figures 5.4a to 5.4e illustrate the classification agreement for the 72 children for whom a comprehensive data set containing a self-assessment by means of the screener (test method), three-day food record (reference method one) and parental completion of screener (reference method two) without any missing values was available.

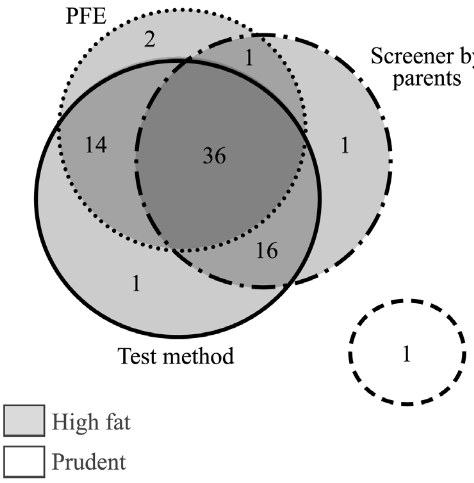


FIGURE 5.4a AGREEMENT: TEST METHOD, PERCENT FAT ENERGY (**PFE**), AND SCREENER BY PARENTS

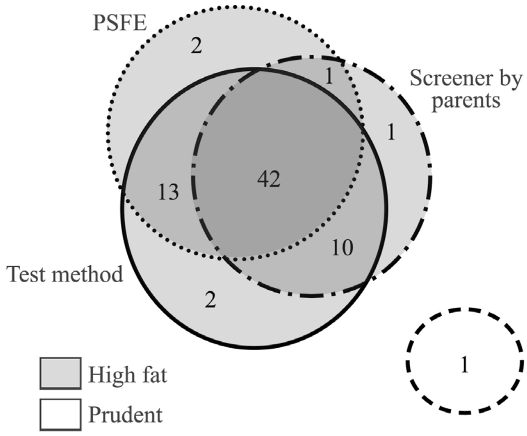


FIGURE 5.4b AGREEMENT: TEST METHOD, PERCENT SATURATED FAT ENERGY (**PSFE**), AND SCREENER BY PARENTS

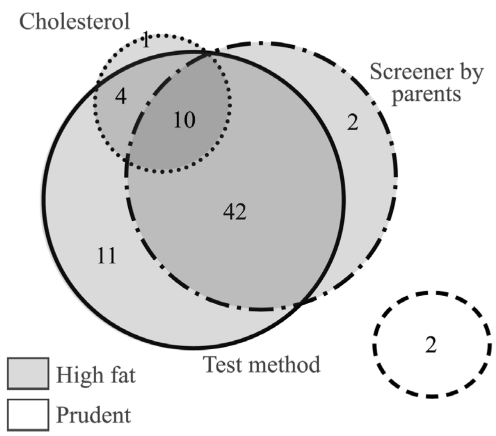


FIGURE 5.4c AGREEMENT: TEST METHOD, **CHOLESTEROL** INTAKE, AND SCREENER BY PARENTS

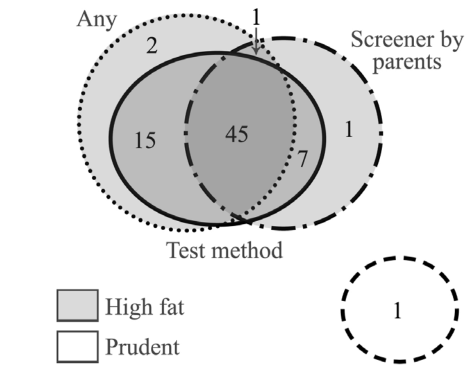


FIGURE 5.4d AGREEMENT: TEST METHOD, ANY FAT MEASURE OF FOOD RECORD (**ANY**), AND SCREENER BY PARENTS

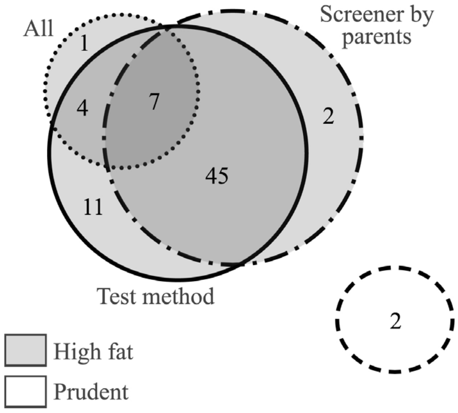


FIGURE 5.4e AGREEMENT: TEST METHOD, ALL FAT MEASURES OF FOOD RECORD (**ALL**), AND SCREENER BY PARENTS

FIGURE 5.4: AGREEMENT AMONG TEST METHOD, FOOD RECORD AND SCREENER BY PARENTS (n=72)

From Figure 5.4 the following is evident:

The percentage perfect agreement between the test method and parental completion of the screener was 73.6% (52 high fat plus 1 prudent out of a total of 72). The number and percentage of participants with identical classifications based on the food record compared to the test method or the parental completion of the screener can be seen in Table 5.19.

TABLE 5.19: IDENTICAL CLASSIFICATIONS BETWEEN FOOD RECORD AND TEST METHOD OR SCREENER BY PARENTS (n=72)

	Food record				
	PFE n (%)	**PSFE** n (%)	**Cholesterol** n (%)	**ANY** n (%)	**ALL** n (%)
Test method	51(70.8)	56(77.8)	16(22.2)	61(84.7)	13(18.1)
Screener by parents	38(52.7)	44(61.1)	12(16.7)	47(65.3)	9(12.5)

There was always a higher percentage exact classification matches between the test method and the food record, than between the parental completion of the screener and the food record. Either way, the highest agreement was found if any of the measures of high fat intake acted as criterion.

The percentage perfect agreement between the screener (either child or parental completion) and the food record was determined by the presence of dietary cholesterol intake as measure. Thus, for Figures 5.4c and 5.4e (which referred to or included cholesterol as outcome measure), the percentage perfect agreement ranged from 12.5% (seven high fat plus two prudent out of 72) to 20.8% (ten high fat plus two prudent), whilst in the absence of cholesterol (Figures 5.4a, 5.4b and 5.4d) the percentage perfect agreement ranged from 52.7% (37 high fat plus one prudent out of 72) to 84.7% (60 high fat plus 1 prudent out of 72).

It follows that perfect agreement within any of the triangulations was also a function of the presence of cholesterol and, as expected, the highest percentage perfect agreement between the test method and both reference methods occurred in 'ANY' (63.8%, that is 45 high fat plus one prudent, Figure 5.4d) even though PFE and PSFE also had considerable perfect agreement among the three methods.

5.6 RECEIVER OPERATING CHARACTERISTICS (ROC)

Figures 5.5a to 5.5e are plots of ROC curves with PFE, PSFE, cholesterol, 'any' and 'all' respectively as standard. The areas under the curve ranged from 0.545 for PFE, 0.548 for cholesterol, 0.555 for 'any', 0.604 for 'all' to 0.654 for PSFE as outcome measure.

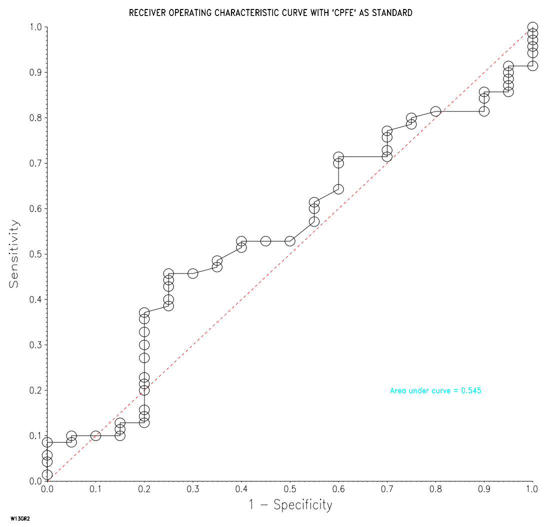


FIG 5.5a PERCENT FAT ENERGY

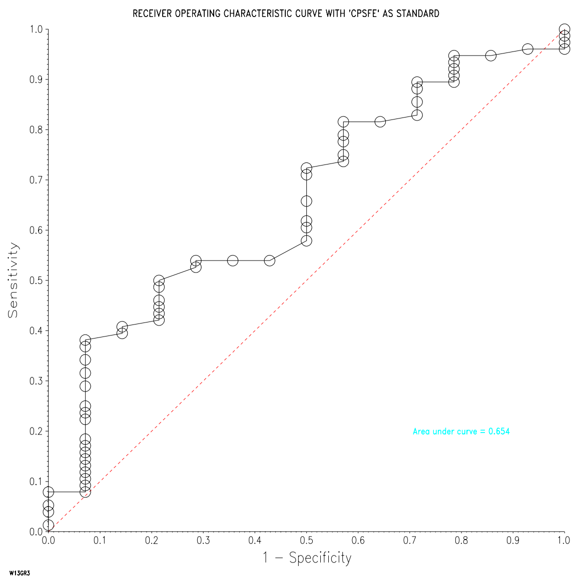


FIG 5.5b PERCENT SATURATED FAT ENERGY

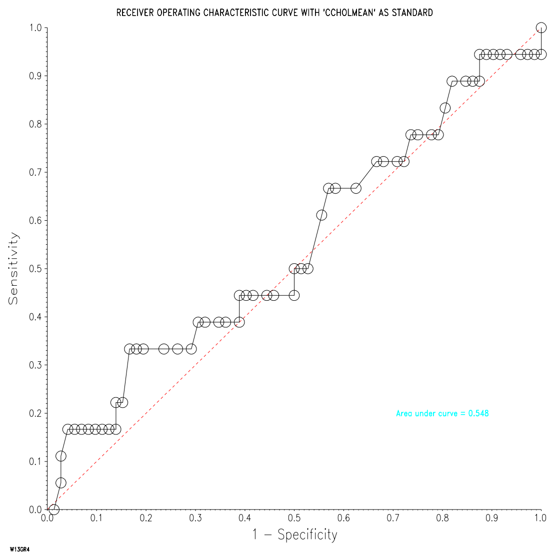


FIG 5.5c CHOLESTEROL INTAKE

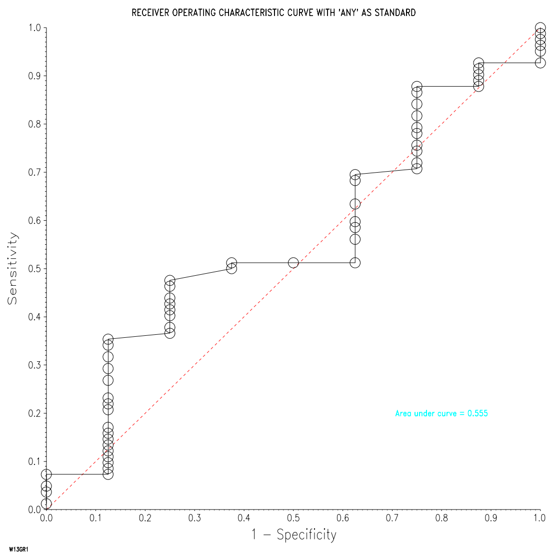


FIG 5.5d 'ANY' MEASURE OF FAT INTAKE

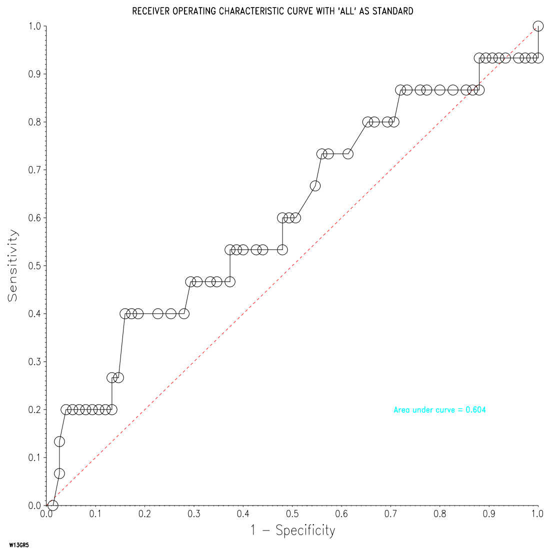


FIG 5.5 e 'ALL' MEASURES OF FAT INTAKE

FIGURE 5.5: RECEIVER OPERATING CHARACTERISTIC CURVES

From the curves it is possible to see how the sensitivity (true positive rate, on Y-axis) and the false positive rate (1 – specificity, on X-axis) co-varied when the cut-off point of the test method was changed. Over the five curves, the sum of sensitivity and specificity was highest at cut-offs of 98 and 118. The effect that changing the cut-off value of the final score of the test method to 98 and 118 would have on the sensitivity and the specificity relative to each of the five outcome measures from the food record is indicated in Table 5.20.

TABLE 5.20: SENSITIVITY (Sens) AND SPECIFICITY (Spec) OF DIFFERENT CUT-OFF VALUES OF THE TEST METHOD RELATIVE TO FOOD RECORD FAT MEASURES OF FAT INTAKE

Test method		**Food record measures of fat intake**									
		PFE [a]		**PSFE** [b]		**Cholesterol**		**ANY** [c]		**ALL** [d]	
		Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
Cut-off value	68 (original)	0.92	0.05	0.95	0.20	0.95	0.08	0.93	0.11	0.94	0.08
	98	0.71	0.40	0.72	0.50	0.72	0.32	0.70	0.37	0.80	0.33
	118	0.46	0.70	0.45	0.79	0.44	0.58	0.44	0.75	0.53	0.60

[a] Percentage fat energy
[b] Percentage saturated fatty acid energy
[c] Any one of the measures of fat intake
[d] All three measures of fat intake

With a cut-off of 98 the sensitivity dropped from the very high values (also reported in Table 5.14) to 0.70 to 0.80 for 'any' and 'all' respectively as standard, but the specificity rose to values ranging from 0.32 (cholesterol) to 0.50 (PSFE). Thus, whilst the sum of the sensitivity and specificity was about the same at cut-off 98 and cut-off 118, the former resulted in a higher sensitivity (true positive rate) and in the latter case it was in favour of specificity (true negative rate).

CHAPTER 6:DISCUSSION

Chapter overview

6.1 DEVELOPMENT AND DEVELOPMENTAL EVALUATION SUB-STUDIES

From the developmental evaluation sub-studies (Chapter 3) the cognitive and other challenges involved when children complete FFQ type dietary assessments became evident. Thus the screener was once more adjusted to accommodate the findings as far as possible without jeopardizing its inherent aims and characteristics, realizing that many issues were still unresolved.

In spite of the attempt to ascertain the content and face validity of the item list by initially checking it in a consensus workshop of experts and then field-testing it in sub-study 1 (face and content validity) it was realised that the list should ideally be based on a recent survey in a group representative of the target population. This may, however, still not be a guarantee of validity: Caan et al [49] recommended this approach to improve the performance of a dietary fat screener that had, in fact, originally been developed by Block et al [44] using a data-base approach. Thus, the developmental evaluation resulted in a tool that was as well adapted as possible for comparative validation in the target group, even though neither the item list, nor the quantifications (reference portion size, frequency of intake categories and the appropriateness of the nutrient database) were claimed to be beyond debate.

Developmental evaluation sub-study 5 (food record) showed that it was feasible to integrate the task of keeping a food diary into the mathematics curriculum and it resulted in well-organised data collection in the main study.

6.2 MAIN STUDY: SAMPLE

Selection bias has been shown to be an important source of error in dietary surveys in the general population [273] and also in children.[226] In respect of the reliability study of this project (referring to measurement of internal consistency and test-retest reproducibility), this was largely ruled out, firstly, because the re-test sample was chosen randomly and based on the absence of a significant difference in final scores (in the first administration) between repeaters and non-repeaters (Table 5.3), and, secondly, because all children were included in the internal consistency part of the study with no drop-outs.

In the comparative validation part of the study the response rate for the food record was good (96%), primarily because of the mathematics (school) context in which it took place. This is in contrast to many studies where poor compliance and a suspicion of non-response bias have been raised with the use of food records. The act of recording could, however, have altered the children's eating habits, a known limitation of food records. For the second reference method (parental completion of screener) the response rate was 72%, which was also considered reasonable. Overall the study was characterised by very few missing values, in contrast to some previous studies involving school children (for example reference [265]), but the external validity was limited by the fact that only one school was included in this stage of the project. Restricting the research to one grade level taught by the same mathematics teacher avoided potentially confounding (intellectual) developmental and administration factors.

Anthropometrically, on average, the participants exceeded median reference indices (weight, height and BMI for age) using the CDC 2000 growth charts as basis. The international cut-off points corresponding to 25kg/m^2 at age 18 for 12.5 year olds are 21.6kg/m^2 and 22.1kg/m^2 for boys and girls respectively.[274] Based on this, the participants' mean BMI's of 20.4kg/m^2 and 20.6kg/m^2 for males and females respectively, were interpreted as being in the 'healthy' range. Thus, in general, the sample could be taken as being anthropmetrically reflective of a population of healthy children.

6.3 TEST METHOD

6.3.1 Internal consistency

Internal reliability, also called homogeneity or uni-dimensionality, reflects the extent to which individual items in a test measure similar characteristics.[231] Therefore it has been reasoned that variance among scores in an internally reliable instrument indicates subject differences and not error.[245]

Keller et al [184] suggested that items with a corrected item-total correlation <0.2 are less relevant for measuring the construct of interest. Since the lowest item-total correlation in this study was 0.35 (Table 5.2), it was concluded that none of the food categories needed to be discarded or rephrased. Cronbach's coefficient alpha values below 0.7 indicate an excess of nuisance items or too few items in a scale. Values >0.7 indicate there are items measuring essentially the same thing and that some are unnecessary.[184] Again it was concluded that all original MEDFICTS items could be retained and that they fully covered the construct under investigation, namely fat intake. The alpha obtained in this study was higher than those included in the review by Yaroch et al.[200]

Several researchers (for example references [233, 236, 246]) have used item total correlations as an indication of convergent validity (a form of construct validity) in nutrition research. In line with this reasoning, it could be concluded that the test method, by being internally consistent, also exhibited convergent validity.

Considering that scale consistency in classical test theory is a function of the number of items in the scale,[231] the various coefficients of internal consistency obtained for this ten-item tool are encouraging for the test method as a screener.

6.3.2 Test-retest reproducibility

The second form of reliability investigated in this study was repeatability or stability over time. If the tool is used with its original aim in mind (that is to assign individuals to the three dietary 'Steps' of the American Heart Association [203, 205]) then reproducibility of this classification is important. In this study the step classification was reduced to 'high fat' and 'prudent' as children in the general population were targeted. In over 90% of cases the first and second administration resulted in an identical classification. It was thus concluded that the screener exhibited classification reproducibility, which would be the prime concern in clinical decision-making.

If, on the other hand, the screener is used for comparing, ranking or monitoring groups of individuals, then reproducibility of the final scores might be a more useful basis for describing reproducibility. Positive correlation coefficients ('reproducibility correlations') between two administrations of dietary assessment tools have often been used as indicative of test-retest reproducibility.[275] Following this line of reasoning, one could conclude that the dietary fat screener exhibited limited ($r=0.36$), yet statistically significant ($P=0.02$) reproducibility in grade six learners as a whole.

The correlation coefficient obtained in this study is substantially lower than those reported for most nutrients by Anderson et al [276] who used a comprehensive quantitative FFQ questionnaire in Norwegian adolescents with mean age 17 years, and for the foods reported by Metcalf et al [277], who had a very heterogeneous group (one to 14 years from different ethnic backgrounds) with a mix of parent, caregiver or child as data source and a reference intake period of four weeks. Also, Andersen et al [251] found (among Norwegian 6th graders) consistent responses ($r=0.62$ to 0.83) on a 24-h recall and a FFQ both re-administered 14 days apart when the previous day's fruit and vegetable intake had to be recorded. By contrast, Yaroch et al [278] administered a picture sort FFQ questionnaire twice during a two-week period to low-income, overweight, African American adolescents, and reported correlation coefficients ranging from 0.28 to 0.36. The reproducibility coefficient of the Youth and Adolescent Questionnaire ranged from 0.26 to 0.58 for different nutrients, and from 0.39 to 0.57 for foods.[279]

Thus, based on reproducibility correlations between final scores, the findings in this study are similar to some and different to other test-retest studies in children, but the obvious differences in research contexts, designs and outcome measures used in the analyses must be kept in mind. The range of correlations for FFQ's in school children reviewed by McPherson [25] was from -0.06 to 0.91, complicating generalisations. Thus, the warnings by Altman [280] and Bellach [275] that using a correlation coefficient to measure agreement may be a "misconceived" or "archaic" analysis, seem appropriate.

Looking at several statistics and exploring the distribution and structure of the measurement error has consequently often been recommended.[111, 154, 275]

The finding that the reproducibility coefficient was non-significant (P=0.29) for boys, whilst it was highly significant (P=0.008) for girls, illustrates that the measurement error may not have been equally distributed amongst participants.

The final score in the screener was the sum of the ten category scores, which, in turn, were the product of the scored portion size and frequency of intake estimations. Thus, focusing on the test-retest reproducibility of the latter two, would further explore the error structure and explain the final score reproducibility. As shown above (Table 5.4), for portion size and weekly intake the percentage agreement was over 50% in eight of the ten food categories. However, when corrected for chance (kappa statistic) the agreement was poor for almost half of the food categories, but overall, the non-agreement tended to be symmetrical. This means that about equal proportions of children changed from a higher to a lower estimate and vice versa.

The kappa's in this study are lower than those reported by Smith et al [255] where middle school students completed a 40-item checklist of foods high in total fat, saturated fat and sodium twice on the same day. These researchers reported kappa values ranging from 0.66 to 0.93. The short period between their administrations may explain the discrepancy. Jonsson et al [267] reported amongst school children 'good' percentage agreement (between 58% and 86%) of usual choice of four different breakfast foods on two occasions eight weeks apart. Based on different statistical analyses, they explain the cases of unreliability in terms of a combination of random and systematic error. As will be evident from the following discussions, this may also be the case in this study.

It is tempting to deduce that methods agree because they are not significantly different. This approach to establishing test-retest reliability has been used in a number of reproducibility studies. Examples include Burden et al [281] (even though they did not report the statistical significance of their findings) and Anderson et al, [276] who found that the first measurement gave significantly higher values than the second. Cullen et al [282] as well as Buzzard et al [265] also reported higher mean consumption estimates in the first administration of their respective screeners. According to McPherson et al [25] this appears to be a trend for FFQ type dietary assessment in school children. Thus in this respect the present study's findings differ from previous reports. The mean difference between the first and second administration for girls was higher than for the group as a whole (2.95±25.42 compared to 0.69±32.59; Table 5.5), but also this value did not significantly differ from zero.

Whilst the small mean difference in final scores (0.69 see Table 5.5) indicates that the two administrations of the dietary fat screener in this study agreed well *on average*, the measures of variability (for example the standard deviation of the difference of 32.6 and the 95% confidence interval ranging from -9.9 to 11.3 for the difference in final score) suggest that for *an individual* absolute agreement of the final scores was less likely. This is confirmed by the Bland-Altman plot (Figure 5.1b) and non-agreement was particularly true for boys. In the case of girls there tended to be less variability (less random error), but a bias (systematic error) towards lower scores in the second administration.

6.3.2.1 Factors affecting reproducibility

Respondent and methodological factors can affect reproducibility.[80, 154, 220] Both of these could have been at work in this study.

6.3.2.1.1 Respondent factors

In the case of children, their cognitive abilities[90], specifically to record, remember or generalise their intake [29] as well as their restricted knowledge of food and food preparation [30] and limited motivation and attention span [81, 283] are well-documented child-specific respondent factors, which can contribute to error. However, true inter- and intra-individual variability also affect the measured reproducibility, for example, in five to 17 year old children the ratio of intra:inter subject variances in intake is, in general, approximately twice that observed in adults [81] and specifically fat intake and fat practices have been shown to vary by meal and day in grade four to six school children.[284] Gender has sometimes emerged as a differentiating respondent factor,[277, 279, 285] but results are conflicting. Age, obesity and weight consciousness are additional respondent factors related to valid dietary assessment.[81, 283, 286] This cannot necessarily be traced to lack of reproducibility, since data may be reproducible, yet invalid. For example, Frank et al [285] found amongst twelve to17 year olds that age did not influence test-retest agreement, yet Bandini et al [287] found that from age ten to15 years girls tended to report energy intake less accurately. The group of children in the present study was very homogeneous in terms of age, educational level and culture. The check whether BMI was related to differences in final scores in the current data set revealed no significant correlation between BMI for age Z-score and difference in final scores in the two administrations, neither for the group as a whole, nor for the genders separately. Thus the fact that in this study the boys in the re-test sample had a BMI for age Z-score of 0.64 (compared to 0.25 for girls, see Table 5.1) did not explain the poorer reproducibility in the boys as sub-group.

6.3.2.1.2 Methodological factors

Amongst the methodological factors in a reproducibility study, the period between administrations remains controversial. Whilst too short intervals will result in learning, carry-over, or recall effects, true changes may occur if the period is too long. Frank et al [285] included a two-hour and two-week interim period in their test-retest reproducibility study in children and found lower agreement in the two-week period. They concluded that a two-week repeat measure tests variability within an individual's eating pattern, rather than the reliability of the instrument. However, most researchers recommend periods from four to eight weeks when assessment of usual diet is the aim.[80] Thus, the higher correlations typically found with briefer intervals,[25] should be interpreted with this in mind. The fact that between the two administrations in this study food recording was done for the validation study, could have affected the second administration. The food categories included in this screener were not so much prone to seasonal variation as fruits and vegetables, which have been highlighted by Joachim [288] as important factors affecting reproducibility.

The type and design of a dietary assessment tool affects reproducibility. Hoelscher et al [289] and Buzzard et al [265] found that composite food items (for example consisting of numerous foods) had poorer reproducibility than single food items. Since all items in this study were composites, this could have affected reproducibility coefficients. The scoring system is another methodological aspect that could have affected reliability. The fact that the screener permitted quite a bit of variability (because portion size had to be specified - in contrast to qualitative and semi-quantitative FFQ - and the scoring system as such, particularly in respect of weekly intake and the multiplication principle) could have reduced reproducibility of the final scores. On the other hand, the classification of final scores into the two classes of fat intake might have increased test-retest agreement. The testing, standardised data-collection and quality control during coding minimise these two factors as strong contributors to apparent poor test-retest reproducibility in terms of final scores in this study. The statistical techniques used and the outcome variables, on which results are based, are analytical factors affecting conclusions in reliability studies, as also evident from the findings of this study.

Reproducibility is a function of sample, geographical and time factors [221], thus again limiting the generalisability of the findings of this study. In addition, precision of differences observed (Table 5.7) is also influenced by sample size. Thus, the relatively small sample size (n=39) also explains some of the variability and consequently the limited reproducibility.

6.3.3 Reflection

Some of the observations made during the administration of the screener are summarised in Table 6.1 in terms of perceived strengths and challenges.

For some food categories, the food list of the test method relied on the ability of the children to differentiate between high fat and low fat versions of outwardly similar foods, for example, for milk and table fats (different types of [tub] margarine). It may well have been that the children were unsure about the type usually eaten, even though the pictures proved very helpful. Thompson et al [118] have recommended that, when different forms of a food exist, it might be helpful to first ask about consumption of the whole food, for example milk, and then the proportion of times each form is consumed (that is 'nesting'). In the present study the filter question "Do you eat foods such as those on the picture?" was intended to fulfill a similar function.

The grouping of items has previously been identified as a FFQ design issue in the sense that multiple, separate questions appear to result in greater accuracy.[118] In screening the aim is rapid assessment, which essentially means losing detail. Ideally this should not be at the expense of accuracy, but in reality it would mean striking a balance. On the other hand, earlier publications have reported that lengthy FFQ's may overestimate intake.[268]

Initially (typically for the first two food categories) the children needed considerable time and very clear, repeated instructions to code their responses on the answer sheets. Then the process speeded up. It was considered to add pictures to the answer sheet, but following consultation with the teacher this was not done, because then learners could potentially work at their own pace, without being briefed about the distinguishing features of the category, thus affecting validity and reliability, and disrupting orderliness.

TABLE 6.1: STRENGTHS AND CHALLENGES OF PRACTICALITIES OF SCREENER ADMINISTRATION

	Strengths	**Challenges**
Venue	• First impression of a 'special' occasion: The conference room was not children's usual domain and evoked comments (for example the comfortable chairs)	• Artificial; not realistic
Group size	• Even though primarily determined by practical constraints (that is size of conference room), group size proved to be ideal from the research perspective	• Each class was divided into three groups, with implications for responsible time management
Setting within venue (U-shape with separators)	• Interviewer could unobtrusively check coding, ensure sustained participation, keep eye-contact, and reduce omissions • Not school-like • Privacy appreciated by children	• Repeatability in other school contexts
Method of administration (Interviewer-guided and-paced; demonstration of coding)	• Data quality • Orderly approach appreciated by school administrators • Can stick to available time (one school period per administration) • Consistency • The visit by a 'dietitian from the university' added status to the children's involvement	• Different work pace of different children (see below discussion)
Administration in school time	• High participation rate	• Can be perceived as disruptive, unless meaningfully integrated into school programme • Requires commitment from involved teachers
Appearance of tool	• Colour and real life pictures were appreciated (visual appeal [290]) • Concrete and realistic • Spontaneous reaction to the pictures. This included positive and negative comments (for example for snacks and organ meat respectively)	• Not all brands pictured • Brands change • Cost • Comments take time and can result in peer influence

Among the practical problems observed with the screener as a FFQ type assessment tool are the following:

- In spite of the example given, facial expressions sometimes suggested that the cognitive challenges associated with reporting frequency of intake were real.
- The children needed repeated reassurance about how to code very low consumption frequencies. This observation may partially explain the general finding that FFQ's tend to overestimate intake.[25]
- Some children found it difficult to separate frequency of intake from a particular portion size, for example milk consumed as drink, on cereals and in coffee / tea. This appears to be different from Thompson et al [118] who claimed that, in general, asking about frequency of intake and about portion size versus frequency of intake of a 'standard' portion size made no difference.
- Linking the word 'medium' to the reference portion size might have implied 'average' or 'normal' or 'recommended'to some children.

In general, many of the cognitive challenges involved in responding to FFQ's as described by Subar et al [117] for adults, were observed during the administration of the test method, in this case to children.

Raat et al [291] used response rate and missing answers as indicators of feasibility, when they tested and validated the Child Health Questionnaire. If these criteria are applied to the current study, it can be concluded that the feasibility was high since the response rate was very high with almost no missing values. The physical setting (that is school context and venue) and the data collection approach (small group, structured interview and coding with teaching aids) probably greatly contributed to this outcome.

All data collection regarding the dietary fat screener had been done by the researcher, a registered dietitian, personally. The characteristics of screening stipulate that a screening tool should be administrable by any qualified professional (that is individuals who are qualified by virtue of their education, experience, competence, or privileges). Thus, whilst the approach ensured consistent administration, it limits conclusions regarding generalisability to other and different health care workers. Inter-rater reliability consequently remains to be established. The standardised text should, however, be helpful in this regard.

The reference period for the test method was since the beginning of the year (that is "since you were in grade six"). This was reflective of about nine months of the year and was assumed to reflect usual intake, since fat intakes are not so much prone to seasonal variations. Furthermore, it was assumed that for South African school children this clear and explicit time frame would be meaningful as it coincided with a cognitive reality, that is the school year. It is, nevertheless, realised that Wolfe et al [239] found that such a relatively long time frame may cause problems for participants.

The fact that trouble was taken to highlight distinguishing feature(s) of each food category helped to avoid interpretation problems and ensure that the grouping of items was clear to the participants, as recommended by Wolfe et al [239] and Livingstone and Robson.[81]

6.4 REFERENCE METHODS

6.4.1 Food record

6.4.1.1 Plausibility of energy intake data

Since the three-day food record was chosen as primary reference method and because it was the only quantitative reference method in this study, it was considered important to establish its plausibility. Reported energy intake has often been used as surrogate indicator of the total quantity of food intake.

In contrast to micronutrients, there are no biochemical markers of energy intake. Three methods of validation are currently available in respect of energy intake data: [159]

- Comparison of self-reported energy intake with the energy intake required to maintain weight.
- Direct comparison of reported energy intake and measured energy expenditure (for example the doubly labeled water technique).
- Comparison of reported energy intake with presumed energy requirements, both expressed as multiples of basal metabolic rate.

Technical and cost considerations excluded the first two methods for use in this study. Consequently the latter method, developed by Goldberg et al,[46] was employed. This consists of the so-called cut-off 1, which tests whether reported energy intake can be representative of long-term habitual intake, and cut-off 2, which may establish whether reported energy intake is a plausible measure of the actual diet during the measurement period. In the Goldberg method the reported energy intake is judged against presumed energy requirements by expressing the energy intake as a multiple of the estimated basic metabolic rate (BMR). The ratio is referred to as the physical activity level (PAL).[46, 159]

Although the concept was originally developed for adults, the principle has been applied in studies with children, for example Torun et al [292] and O'Connor et al.[293] A PAL of 1.06 has been used as cut-off to test whether reported energy intake from a three-day record is a plausible measure of the food consumed during the actual measurement of dietary intake. [43, 293] Nevertheless, Livingstone et al [294] recently cautioned against its application for identifying individual misreporters in paediatric groups.

Consequently the PAL was used in this study to establish whether overall bias at group level was present and not to discard the twelve individual records below 1.06. The following additional, inter-related reasons are presented for this decision:

- It has been claimed that excluding underreporters may introduce an unknown bias into the data set,[159] because underreporting is the result of undereating (eating less whilst recording) plus underrecording (failure to record everything that was eaten).[295]
- Omissions and intrusions have been found to result in low accuracy and low consistency in grade four children's dietary recalls,[296] but misreporting (under- and/or overreporting) may also be selective in terms of certain nutrients or foods.[56, 243]
- Underreporting appears to be not equally distributed within a population. Weight status, sociodemographic and psycho-behavioural factors may all be predictors of underreporting [159, 230] and contribute to the so-called subject-specific bias which seems to be characteristic of specific individuals regardless of the dietary assessment method used and persistent over time,[297] as discussed in the review of literature (Figure 2.3).
- Finally, after critically assessing the sensitivity and specificity of the original Goldberg cut-off, Black [298] concluded that information of each subject's activity level is also necessary to identify diet reports of poor validity.

The mean PAL of 1.45±0.39 found in this study thus suggested that, on average, the energy intake was plausible for the reporting period. The value, however, was lower than those related to light habitual physical activity, namely 1.54 and 1.48 for six to 13 year old boys and girls respectively.[292] A PAL >=1.4 but <1.6 for boys and girls between ages nine and 18 (within the 5th to 85th percentile for BMI) corresponds to 'low active' in the four physical levels published by the Institute of Medicine.[272]

A bias in dietary reporting is thus probable also in study, similar to previous studies in children [25, 81, 159, 299, 300] where often doubly labeled water acted as reference. Thus, the energy intake could not be regarded as representing long-term habitual diets.

Underreporting may be intentional and unintentional. In general, food recording can be subject to problems like changes in intake to simplify recording, decisions not to eat or record foods that might be considered 'undesirable' and failure to record food that were eaten in excessive amounts.[243]

Factors that could, in addition, have contributed to the underreporting specifically in this study include the following:

- A tendency toward recording fatigue, evident in consecutive day recording, could be noticed as, overall, mean energy intake on the third day was significantly lower than on the first day of recording (Table 5.9).
- Weekend days were not proportionally represented and it was found that energy intakes were always higher on weekend days compared to the other days of recording of a particular recording group, regardless of whether the weekend day was the first or the last day of recording. For the two recording periods that included a weekend day no recording fatigue was shown. This was surprising for the Sunday to Tuesday group. Thus it could also be that during weekends parents became involved and affected the recording 'pattern'.
- Whilst the use of the electronic scales was shown during the developmental evaluation sub-study to be associated with more comprehensive *descriptions* of food consumed, the results in the main study indicate that the inconvenience of weighing (in contrast to estimating) may have resulted in underrecording and / or undereating, because the children weighing their food recorded lower intakes than those using household measures. Almost double the number of children weighed their foods, thus affecting the mean energy intakes.
- The fact that for girls a non-significant correlation between weight and mean energy intakes was found, confirmed the presence of selective (gender-specific) personal bias,[292] since higher habitual energy intakes can be assumed for persons with higher weights. Based on the PAL, the percentage of DRI for energy and the correlations between body weight and reported energy intakes, it thus appeared that the boys' food records were more plausible than the those of the girls.

- The BMI z-score of the girls in this sample (0.43±1.1) suggests that these girls' BMI's were, on average, higher than the reference population, also a variable known to be associated with underreporting.[297, 299]

It follows that a coalescence of methodological (that is food record-related) and respondent (personal) factors seemed to have played a role in the underreporting. This is in line with previous studies on the characterisation of low energy reporting for example Cook et al [301] and as indicated in Figure 2.3.

The observed day-to-day variability in intakes of energy, fat, saturated fat and cholesterol relative to international standards, that is standard deviations and coefficients of variability from then Continuing Survey of Food Intakes by Individuals,[272] suggested overall comparability of this study to other publications.

6.4.1.2 Reflection

Whilst the strengths of the food record (see review of literature) explain why it is generally used as reference method, the drawbacks limit its use. In this study an attempt was made to administer the food record in such a way to minimise some of the weaknesses without losing the strengths. The following was noted in respect of the approach and context used:

Firstly, the respondent burden associated with completing food records is well known. By integrating the food recording into mathematics, the process became a school assignment and not a task to be completed over and above the daily chores of learner and teacher. Nevertheless, commitment from the responsible teacher remained crucial. Helping the teacher set up meaningful follow-up assignments was a form of recognition and an attempt to contribute to curriculum development. The current paradigm shift within the South African education system could thus be used to create a win-win situation for nutrition research and the involved school. It is realised that the context as a whole may have created some pressure for the individual learner to participate.

Secondly, the fact that all learners were included addressed selection bias, which is very often associated with food records, because those who complete the records may be significantly different from those who do not participate.

Thirdly, since the whole class was involved, the act of recording did not make the participant feel 'out' when weighing and recording foods. It remained, however, important to maintain the motivation and ensure that (negative) peer pressure resulted in not (properly) doing the task at hand, overdoing it or in terms of changing usual eating habits. Again the daily reinforcement and guidance by the teacher was critical; also for checking that weighing and recording techniques were adequately performed.

Providing scales and household measures added value to the mathematics learning experience, in addition to stimulating the children. The logistics, detailed learner training, briefing and daily re-briefing were perceived as time consuming, particularly when the food recording extended beyond one day and the learning benefit related to mathematics decreased.

Employing a dietitian of the same culture as the children, who was up to date with their eating habits, language usage and trends in the food industry, proved to be very valuable, since the handwritten records showed great variations in the participants' language usage (terminology used and spelling abilities, for example "kaaikiebotter" [peanut butter], "myjenys" [mayonnaise], "niekerball" [sweet], "stuee" [stew)]) and participants often recorded foods by their brand names.

Including measures of quality control (in terms of anthropometric and design factors) proved valuable for explaining results. Future investigators could consider refinement by combining estimated BMR with physical activity measured objectively (for example with a triaxial accelerometer [302]).

In general, whilst for boys the test-retest reproducibility of the test method was shown to be worse than for girls (see previous discussion), the credibility of the food records was better for boys than for girls. This might be explained in terms of gender-related food awareness: In the test method, which relied on memory, the presence of random error suggested that the boys were guessing, whilst the girls tended to be more consistent (but perhaps consistently underreporting). When the task at hand was to record intake, the boys appeared to be closer to the 'truth' and the girls still underrecorded. This implied a gender-specific error structure: For boys poor comparative validity of the test method should thus already at this stage be expected, because of the low reproducibility of the test method, whilst for girls the potential of agreement with the reference method existed. However, there was a good chance that for girls both, the test and the reference method, reflected underreporting, and both would thus be 'untrue': The dilemma of a 'non-golden' reference method.

6.4.2 Screener by parents

Reporting on fat intake requires some knowledge of foods, food preparation and food composition (in terms of brands and food labels) in order to differentiate between, for example, different types of meat, dairy products, table fats and cooking methods, all of which children might not have. Therefore, using parents as surrogate sources of information in the comparative validation was deemed important.

The findings of the comparison between children and parents are discussed in detail as part of the following sections.

6.5 COMPARATIVE VALIDATION

6.5.1 Basic associations between test method and reference methods

The processing involved in the test and reference methods allowed for several approaches and outcome variables to be used in the comparative validation. In the following section the results of the comparison of the test method to the reference methods in respect of the non-classified final outcome are discussed and interpreted. Since no dietary assessment method is perfect, a critical analysis and understanding of the nature and source(s) of the underlying error forms the backbone of comparative validation.

6.5.1.1 Test method versus food record

Against the background of limited reproducibility of the test method amongst boys in this study, the lack of statistically significant correlations between the final score obtained in the screener and the three outcome measures from the three-day food record (PFE, PSFE and cholesterol) was not surprising. Equally, the relatively low (yet statistically significant) test-retest reproducibility coefficient for the group as whole (based on final scores in the test-retest assessment), partly explains the absence of an association between test method and the food record. Thus, since reproducibility is a requisite for validity, only for the girls a meaningful comparative validity could be expected.

The aim of MEDFICTS is to predict high PFE, PSFE and cholesterol intakes. The current study did not yield significant correlations between the final score of the test method and any of these measures of fat intake for the group as a whole. This is in contrast to the results of Srinath et al,

[201] Kris-Etherton et al [203] and Taylor et al.[205] Possible reasons could be the nature of the current target group (age and the effect thereof on both, the test method and the reference method, as well as cultural differences) and/or the modifications made to the tool.

The finding that total fat intakes (rather than energy contributions) revealed significant correlations, has previously been documented.[199, 206, 217] Van Assema et al [199] explained it by very high correlations between total fat intake and total saturated fat intake resulting in a low variance in percentage energy from fat, which could lead to low correlations between test method final scores and measures of fat intake from the food record. A further explanation is probably the fact that the test method did not include non-fat energy sources.

Rohrmann and Klein [264] also used total intakes (as opposed to energy contributions) as outcome measures in their validation study of a dietary screener. They reported correlation coefficients of 0.44, 0.50 and 0.56 ($P<0.001$) for total fat, saturated fatty acids and cholesterol respectively. The items in their short questionnaire were based on a representative food consumption survey of the target population, they did not measure portion size, the target group consisted of adults, and the reference method was a 148-item FFQ, all in contrast to the present study and all of which could offer an explanation for the higher correlation coefficients they obtained.

Caan et al [49] evaluated the performance of a dietary fat screener and reported that it was more effective at classifying respondents into quintiles of total fat intake than into quintiles of percentage of energy from fat.

The limitations of using correlation coefficients to establish the validity of a dietary assessment method have been mentioned before and are well documented.[27, 88, 154, 270] The main problem is that it cannot be judged on a null hypothesis basis. Furthermore the confounding effect of intra-subject variation on usual intakes is not taken into account and consequently, because the reference method in dietary assessment itself is usually imperfect, a correlation coefficient may underestimate the level of agreement (attenuation bias) with the actual usual intake. The reason for reporting correlation coefficients in this study is thus primarily for comparing results to previous research. Furthermore, in a study comparing different statistical methods for assessing relative validity of a FFQ, Spearman correlations were found to be useful [303] and the application of a combination of statistical methods was again highlighted.[304]

PFE and PSFE intakes of children classified by the test method as consuming a high fat diet were very similar to those classified as prudent eaters, but absolute intakes (energy, total and saturated fat, cholesterol) were higher in the high fat group. This suggests lack of concurrent validity,[198] when two of the intended outcomes of the test method, that is PFE and PSFE, were used as standard. As in the case with the correlations discussed above, the test method tended to be stronger related to absolute intakes compared to proportional intakes.

6.5.1.2 Test method versus screener by parents

As indicated in Chapter 5, when the screener was in the hand of the parents in respect of their grade six child, it also exhibited characteristics of homogeneity. The findings from the item total correlations, Cronbach's alpha and the split half method were comparable to what was found when the screener was used by the children themselves to perform a self-assessment of intake (see previous section on test method). It is thus concluded that the dietary fat screener *per se*, regardless of the data source (that is the children as primary informants or the parents as surrogates) was internally reliable, which could be taken as enhancing its content validity.[231] In the original pilot testing of MEDFICTS Srinath et al [202] had already noted that self-administered and interviewer-administered application of the tool resulted in similar findings.

If, on the other hand, the outcomes of the measurement by parents of their children's diets were compared to the outcomes obtained from the children themselves, the following was found: Firstly, when the two primary building blocks of the screener (that is the reported usual portion size and weekly consumption) were analyzed (Table 5.16), it appeared that:

- In general (for both, portion size and weekly consumption, as well as across the food categories) there was limited agreement between parents and children. This was reflected by relatively low percentages of identical responses and also few food categories with fair or moderate chance corrected agreement. In three cases the kappa value was in fact negative, meaning worse than chance agreement. Hoehler [305] has argued that the presence of bias reduces kappa values. As evident from the following discussion there appeared to be systematic error in terms of reported portion size and weekly consumption, thus explaining some of the very poor chance corrected agreement obtained between children and parents.

- There appeared to be a systematic error in the sense that most of the non-agreeing responses were not symmetrical (McNemar data in Table 5.16). Typically parents reported smaller portion sizes and less frequent consumption. This seemed to have had a 'carry-over-effect' to the category and final scores as evident from Table 5.17 and Figures 5.3. The reason for the lower scores by the parents may be either due to underreporting by the parents or overreporting by the children, or both. For preschool children it has been found that mothers were more likely to underreport then over-report foods.[81] In this study the difference appeared to be more evident for the food categories to which the parents may have had an 'unhealthy' connotation (for example dessert, convenience foods, fried and baked foods, as compared to meat, milk and cheese) whilst being popular (and possibly overreported) by the children as suggested by Koehler et al. [92] This does, however, not explain the finding in respect of snacks.

- Whilst reported frequency of intake *and* portion size showed lack of agreement between parents and children, this appeared to be more evident in respect of portion size. The lack of symmetry was equally common in reported portion size and weekly consumption.

- Thus, at least one of the data sources (children or parents) appeared to lack validity in terms of reported portion size and weekly consumption using the dietary fat screener. Matheson et al [306] assessed the validity of eight to twelve year old African American girls' self-report of food portion estimates and found "sizable errors in quantitative estimates". Consequently cautious interpretation of the children's self-reports of portion estimates seems to be necessary. On the other hand, from the food records it was evident that the children made many food choices in the absence of their parents, making parental error also not unlikely.[81, 307]

Secondly, the correlations between parents' and children's individual category scores (which are the product of the scored portion size and weekly consumption) were statistically significant for only three, borderline for two, and non-significant for five of the ten food categories.

Thirdly, the correlation between the sums of the category scores (that is the final scores of parents versus children) was small (r=0.23), yet statistically significant (P=0.04). A differential pattern emerged for boys and girls, with parents and their sons not showing a linear relationship. Since the boys' final scores had also not been reproducible (see previous discussion) this is not surprising, as reproducibility is a prerequisite for validity. In the case of girls, the statistically significant positive correlation coefficient (r=0.33, P=0.04) between the parents and their daughters only shows that higher values in the one group were associated with higher values in the other group.

The finding that the mean difference between parent-daughter pairs' final scores differed significantly from zero confirmed earlier suggestions (see frequency of intake and portion size) of systematic error. This was also evident from the plots of the individual data points (Figures 5.3a and 5.3b), which additionally show the wide scatter around the diagonal (Figure 5.3a) or horizontal zero line (Figure 5.3b) of perfect agreement, suggesting considerable variability for individual pairs.

In spite of a relatively high percentage (76%) of identical classifications into high fat or prudent intake, the chance-corrected agreement between parents'and children's final classifications was poor. According to Hoehler [305] this may have two reasons: Firstly the presence of bias, which was shown to be the case in this study and, secondly, by prevalence effects, which also applied to this study. The latter occurs when the model is based on an underlying continuous variable; in this case the final score.

6.5.2 Classification agreement

In many situations the aim of the test method as screener would be primarily to classify individual intakes into 'high fat' versus 'prudent'. In this case classification agreement between the test method and the reference methods would be of prime interest.

A tool's overall predictive value is defined as its ability to predict correctly the presence or absence of nutritional risk. Consequently the dietary fat screener's overall predictive value refers to its ability to predict correctly the presence or absence of high fat intake (based on references [45, 308]). The percentage perfect classification agreement between the test method on the one hand, and the two reference methods on the other hand, provides an overview of the test method's overall predictive value and thus an indication of criterion-related validity. Figures 5.4a to 5.4e graphically represent the findings from the triangulation and are discussed below (based on n=72), even though the classification agreements between the test method and each of the reference methods separately (based on n=93 and n=78 respectively for reference method 1 and 2) were also presented in the results section.

The screener, when completed by grade six children or their parents, showed some classification agreement (about 74%). Percentage agreement between the test method and food record tended to be similar or slightly higher than this, when PFE, PSFE or any of the three measures of high fat intake acted as criterion (71, 78 and 85% respectively). When cholesterol intake was included as reference, the agreement was much lower (about 20%). The agreement between the two reference methods followed a similar pattern, but was always less than the corresponding agreement between the test method and the food record.

Whilst the three-day food record in this study was shown to be plausible in respect of the recording period, the PAL values obtained suggest that it was probably not reflective of 'usual' energy intake. Consequently it may also have been inadequate in capturing 'usual' cholesterol intake. Furthermore, cholesterol (compared to total fat and saturated fatty acid) intake is known to have a high intra-subject variance for the general population [27] and even more so for children (five to 17 years old).[81] The number of days required, on average, to estimate true usual adult intakes of cholesterol might be as high as 139-200 and 13-15 days for an individual and a group of individuals respectively. [167] The test method and the screener completed by the parents showed poor agreement with the food record classification based on cholesterol intake, yet they tended to agree (in terms of percent identical classifications) with one another. Thus, it is concluded that the three-day food record may have been an inappropriate reference method in this respect, rather than the test method being an inadequate tool for screening for high usual cholesterol intakes.

6.5.2.1 Sensitivity and specificity

Measuring sensitivity and specificity to describe the validity of dichotomous screening tests is very common in the medical literature. [309, 310, 311] It is also increasingly used in nutrition research (for example references [33, 49, 193, 312]).

Even though no dietary assessment method is flawless, the weighed food record is an accepted, practical relative standard.[157] Consequently the criterion-related validity of the test method in terms of the sensitivity and specificity was expressed relative to the various measures of fat intake obtained from the food record (see Table 5.14).

Sensitivity is the ability of the test method to correctly identify individuals truly at nutritional risk, that is true positives. Thus, in the present context it would refer to the dietary fat screener's ability to identify correctly children who, according to the three-day weighed food record, had a high fat intake.

Specificity measures the test method's ability to correctly identify persons who are not at nutritional risk, that is true negatives. In analogy it would refer to the dietary fat screener's ability to identify correctly the children who, based on the three-day food record, consumed a prudent diet.

From Table 5.14 it is evident that for the dietary fat screener with a cut-off value of 68 there would be considerable misclassification of those not at nutritional risk as determined by the food records and to a much lesser extent of those at nutritional risk, because the dietary fat screener exhibited high sensitivity in identifying high fat intakes, but lacked specificity. In the clinical medicine context, a large sensitivity means that a negative test can rule out the disease (David Sacket coined the acronym “SnNOut” for this); thus, for the dietary fat screener with its high sensitivity it could be concluded that a child for whom the result indicated prudent intake, high fat intakes could be ruled out. Equally, a large specificity would have meant that a positive test could rule in high fat intakes (The David Sacket acronym for this: "SpPIn"). The latter was, however, not the case for the dietary fat screener in the current study.

The ideal would be to have both, high sensitivity and high specificity, in a screening tool. It is, nevertheless, well known that in real life usually a balance must be struck between sensitivity and specificity, specifically when the test variable is a continuous variable. Decisions on the appropriate cut-off for a screening test mainly depends on the consequences of identifying false negatives and false positives,[45] but also on the implications of the test for the patient and the health care system and availability of effective treatment.[310] The purpose of a particular study may also play a role, for example if the test method is used to establish prevalence the aim may be to have a balance between false positives and false negatives. In a study that compares rates in different populations the absolute rates may be less important, but the primary concern would be to avoid systematic bias, meaning a specific test may be preferred, even at the price of some loss of sensitivity.

In the context of this study, clearly it would be undesirable to have many false negatives, that is, failure to identify those who are really at risk of high fat intakes, because the benefits of early detection and intervention that are associated with preventing CNCD in childhood [15, 19, 20] would be missed. It has been argued that general screening tools aimed at detecting malnutrition should primarily be sensitive, because an in-depth, follow-up assessment to confirm the screening result is presumed, or because it is reasoned that giving nutritional care to those who do not need it does no harm. Furthermore, early diagnosis generally has intuitive appeal, as earlier treatment is thought to be related to improved prognosis.[310]

The practical and cost implications of unnecessary screening, follow-up and interventions should, however, also be considered, apart from potential harms related to anxiety, adverse effects of labeling (stigmatisation and discrimination), inconvenience and the possibility of childhood risk factors not tracking into adulthood (inconsequential risk).[309, 310, 313] From the above it follows that a high proportion of false positives is also not acceptable.

Wald et al [314] have argued that a risk factor has to be extremely strongly associated with a disease within a population before it can be considered to be a potentially useful screening test. Even a odds ratio of 200 between the highest and lowest fifths will yield a detection rate of no more than about 56% for a 50% false positive rate. Another reason why strong risk factors may make poor screening tests, according to these researchers, is that there may be little variation in exposure within populations. They explain this by referring to the smoking example: It is known that smoking cigarettes is a risk factor for lung cancer. However, if everyone in a certain population smoked 20 cigarettes a day, asking about cigarette consumption would not distinguish those who are more likely to develop lung cancer from those who are not. In the present study the high prevalence of high fat intake could have had a similar effect.

As evident from Table 5.13 relatively few true negatives were found in all the comparisons between the screener and the measures of fat intake from the food record. This would contribute to the wide confidence intervals of the reported specificity (Table 5.14). Again the sample size would have played a role, as the required number of negatives necessary to yield the desirable power (see 4.1.1) was not obtained.

6.5.2.2 Receiver operating characteristic (ROC) curves

The high sensitivity but low specificity observed in this study was the rationale for investigating the effect of changing the cut-off value of the final score of the test method on the sensitivity and specificity relative to the various measures of high fat intake from the food record. By increasing the 'strictness' that is increasing the final score cut-off of the test method, the false positive rate should decrease, the sensitivity would also decrease with an associated increase in specificity.

ROC curves are useful to depict this pattern of sensitivities and specificities observed when the performance of the test method is evaluated at different cut-off values. They thus describe the whole set of (1-specificity, sensitivity), that is (false positive fraction, true positive fraction) combinations possible.[47]

In a ROC curve a test that perfectly discriminates between the two groups under discussion (that is ‘high fat’ versus ‘prudent’ diet consumers) would yield a curve that coincided with the left and top sides of the plot, that is having a high sensitivity (true positive rate) and a low false positive rate (1 minus specificity). Poor tests have lines close to the rising diagonal and consequently the area under the curve would be about 0.5. Hosmer and Lemeshow [315] have suggested that areas of 0.7 and higher can be taken as pointing to acceptable discrimination abilities of the test method. The shape of the curves obtained in this study showed that, regardless of which measure of high fat intake was used, the discrimination ability of the test method remained low, as the highest area under the curve was 0.65 (for PSFE). Equally, it was not possible to optimise the test method by manipulating the cut-off value of the final score.

This finding was in contrast to the Taylor et al study [205] where it was found that by substantially decreasing the cut-off value of the original MEDFICTS tool [203] a sensitivity and specificity of 0.73 and 0.75 respectively could be reached.

The results of the current study were more in line with those of Caan et al.[49] They reported that the sensitivity and specificity of a dietary fat screener varied depending on the cut-off point used, but it was not possible to achieve high sensitivity and high specificity simultaneously. Also Prochaska et al [316] found that their screening measure of fat intake was sensitive but not specific among adolescents.

In the development rationale of the test method (Chapter 3) the primary aim of the test method was stipulated as that it should be able to discriminate between children who consume high fat diets and those with prudent intakes. The ROC curve suggested that in the described context the dietary fat screener did not achieve this eventual goal.

CHAPTER 7:CONCLUSIONS AND RECOMMENDATIONS

"*Ever since records of diet were first kept, we had an unerring but misguided confidence in our ability to measure our own food consumption - until, that is, nutritional epidemiology revealed to us the error of our ways*."

These words of Nelson [154] summarize the essence of this study and confirm that the contribution of a dietary validation study primarily lies in describing the nature and magnitude of the error structure of the measurement of diet. In this regard the present study additionally indicated how the error structure of a FFQ type dietary fat screener might differ for subgroups and different food categories, how it might be affected by the choice of reference method and statistics used and, finally, how the aim (outcome measure) could determine the conclusions. Thus, in the present study the measurement of grade six learners with the dietary fat screener

- was internally consistent
- was reproducible in the girls, but random error characterized the boys' assessment
- did not agree with screening performed by parents, with systematically lower overall reporting by the parents
- had a different error structure for the different food categories in the tool and for the various quantification elements (portion size and frequency of intake and the resultant scores)
- showed selective agreement with measures of high fat intake (PFE, PSFE, and cholesterol intake) from a three-day food record. The agreement depended on the statistical analysis and the outcome measure used for comparison
- was highly sensitive, but could not achieve good specificity simultaneously.

Nutritional exposure can be defined on three different levels of measurement: as food, nutrients and biomarkers.[162] The test method in this study measured on the food level, the primary level of exposure, but it was intended to relate to PFE, PSFE and cholesterol intakes, all on the nutrient level. Thus the research design of this study, where the test-retest reproducibility and screener completed by the parents (reference method 2) addressed the food-level agreement, and the three-day food record (reference method 1) focused on nutrient-level agreement, was considered an appropriate and strong approach, also incorporating the triangulation principle. The inclusion of anthropometric and design quality control specifically with respect to the food record further strengthened the internal validity of the study. Complementing the design with a biomarker, with a completely independent error structure, could have 'perfected' the comparative validation.[317]

The validation process is sometimes considered to relate to the measurement and not to the method from which the measurement is derived,[154] meaning that validation considers the context within which dietary assessment methods are used. Consequently all conclusions derived from this study primarily relate to the given context: A public, urban, middle-class primary school in South Africa, accommodating mainly white, Afrikaans speaking, children. Nevertheless, the application of scientific design principles and quality control measures within the quantitative domain of investigation do allow some generalizations.

The test method is in essence a FFQ. Drewnowski [98] has argued that FFQ estimates do not appear to be based on memory for actual events, but that food frequencies are inferred, as opposed to remembered, and are based on some subjective image of a usual or typical diet. As such, he argues, FFQ's cannot be 'validated' since they measure primarily predispositions and attitudes. They can thus not be compared to instruments that capture actual behaviour in the short term. Even though this is not a commonly held perspective in the nutrition literature, the reasoning does provide some additional explanation to the limited agreement between the test method as a FFQ and the three-day food record in the current study.

A similar note is struck when the question arises whether a 'usual / habitual / typical' diet exists. Is it only a construct in the minds of dietitians and nutritionists or is it an objective entity? Whilst it has been operationally defined (for example the average in a long series of food records [157]) ever since the mentioned study of Huenemann and Turner (see review of literature)[238] the existence of a 'usual' diet has from time to time been debated. Thus, again, should a 'usual diet' not exist, validation is either not possible or, at best, construct validity (in contrast to criterion validity) would be an option.

If, however, a 'usual diet' does exist, then the three-day food record could rightly be criticized as being an inappropriate reference method for validating the test method. Furthermore, given the fact that the three-day food record is an imperfect measure of dietary intake, it would have been ideal had the test-retest reproducibility thereof also been determined. Only then, when the variability (random error structure) of the reference method is also known, can more definitive conclusions about potential relationships be drawn. This, of course, also applies to the screener completed by the parents.

The above has implications for the selection of the sample. The lack of due consideration of sample size for a validation study has been addressed (for example references [185, 166, 318]) but

Keller et al,[184] have suggested that, in general, a sample size of 100 should be adequate. Representativeness and composition of the sample also demands attention: The high prevalence of high fat intakes observed in the present study may be a true reflection of reality in the study group, but it may have reduced variability and may have affected some statistics for example some of the statistical indicators of agreement such as the kappa statistic, are affected by prevalence.[305] By trying to obtain a population with trait prevalence near 50% this could be addressed.[305] However, 'manipulation' of the sample affects the relevance of other indicators, such as positive and negative predictive value.

There are many ways of analyzing and expressing reliability and validation studies [319] and the most appropriate statistical analysis has not been established.[28, 42, 122, 154, 304] This also appears to be true for methods specifically aiming to measure fat intake: Simon et al [320] compared standard methods based on a null hypothesis of no agreement between instruments (FFQ, 24h recall and three-day food record) and an alternative method of analysis based on a null hypothesis that the instruments should be in agreement. They conclude that the latter is more appropriate. Jones [185] reviewed and critically appraised the scientific merit and methodologies used for nutritional screening and assessment tools and concluded, "no one tool is judged to have been published with sufficient care given to its application, development and evaluation." This was confirmed by Dennis et al [147] specifically in respect of the design and reporting of FFQ. Their scoring method and the Consensus Document on the development of FFQ [114] should provide more design and analysis guidance for the future. Close collaboration with bio-statisticians seems to be indicated. In this study a variety of well-established and novel statistical analyses were reported in order to provide a comprehensive picture and enable comparisons to other studies.

In conclusion, the dietary fat screener should not yet be used in grade six learners as a sole assessment method within the South African primary health care context, given the country's present, overall nutrition profile [321] and available health care resources. Screening is an inexact science. For that reason ethical and legal responsibility should rest on those administering it to inform the public of a particular tool's discriminatory properties.[313] The data obtained in this study suggest that if intervention or monitoring of dietary intake trends are to be based on only the dietary fat screener, further developments and / or modifications to increase its validity are needed.

Possibilities for structural changes to the tool include re-scrutinizing the item list (for example critically evaluating the role of eggs as food category or adjusting the relative weight of the individual food items), ideally based on a food consumption survey in a (nationally) representative sample including the target group. Alternatively, statistical modeling of expert judgment matrices could be used to obtain an indication of the relative importance of the individual food categories. Critical investigation of alternative scoring principles is another avenue to investigate. It is furthermore possible that including more covariates (for example BMI or gender) into the logistic regression could result in improved discrimination abilities of the screener. Checking the assumptions of the nutrient data for South Africa could also be useful. Finally, deeper insight into the cognitive processes of dietary assessment of children in the target group could be helpful, but without sacrificing the inherent strengths of a screening tool.

In spite of the identified limitations, given the high prevalence of high fat intakes in the target group (and thus the risk for developing CNCD), the tool may in the interim be very valuable for creating awareness of high fat intakes. The food-based nature of the screener should be a practical starting point for providing needs-driven nutrition education and anticipatory guidance (similar to the approach used with the REAP and WAVE tools),[34] within population-wide promotion of the dietary guidelines.

Once measurement with the dietary fat screener has been shown to be reproducible and valid in this target group, expansion of the target group and context in line with the outline in Figure 1.1 is recommended. In doing so a greater segment of the South African society may eventually benefit from the research. This should be followed up with randomized controlled trails of screening using the designs suggested by Barrett et al [309] in order to ascertain cost-effectiveness of the process in the South African context.

REFERENCES

1. Berrios X, Koponen T, Huiguang T, Khaltaev N, Puska P, Nissinen A. Distribution and prevalence of major risk factors of noncommunicable diseases in selected countries: The WHO Inter-health programme. Bull Wld Health Org 1997;75:99-108.
2. Steyn K, Fourie J, Bradshaw D. The impact of chronic diseases of lifestyle and their major risk factors on mortality in South Africa. SA Med J 1992;82:227-31.
3. Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease study. Lancet 1997;349:1269-76.
4. Murray CJL, Lopez AD. Alternative projections of mortality and disability by causes 1990-2020. Lancet 1997;349:1498-1504.
5. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-1025. Diabetes Care 1998;21(9):1414-31.
6. Bellamy C. The state of the world children 2003. New York: UNICEF; 2002.
7. Drewnowski A, Popkin BM. The nutrition transition: New trends in the global diet. Nutr Rev 1997;55(2):31-43.
8. Walker ARP. Nutrition-related disease in Southern Africa: with special reference to urban African populations in transition. Nutr Res 1995;15(7):1053-94.
9. Lucas A. Programming by early nutrition - an experimental approach. J Nutr 1998;128(2 suppl):S401-6.
10. Mahan LK, Escott-Stump S. Krause's food, nutrition, and diet therapy. 11th ed. Philadelphia: Saunders; 2004.
11. Krauss RM, Eckel RH, Howard B, Appel LJ, Daniels SR, Deckelbaum RJ et al. Revision 2000: A statement for healthcare professionals from the nutrition committee of the American Heart Association. J Nutr 2001;131:132-46.
12. Ballard-Barbash R. Designing surveillance systems to address emerging issues in diet and health. J Nutr 2001;131(2SI):437S-9S.
13. Vorster HH, Love P, Browne C. Development of food-based dietary guidelines for South Africa - the process. SA J Clin Nutr 2001;14(suppl)(3):S3-6.
14. Love PV. Development of food-based dietary guidelines for infants and young children living with HIV/AIDS. SA J Clin Nutr 2004;17(suppl)(2):S45.
15. Kavey R-E, Daniels SR, Lauer RM, Atkins DL, Hayman LL, Taubert K. American Heart Association guidelines for primary prevention of artherosclerotic cardiovascular disease beginning in childhood. Circulation 2003;107:1562-6.
16. American Dietetic Association. Position of the American Dietetic Association: Dietary guidance for healthy children ages 2 to 11 years. J Am Diet Assoc 2004;104:660-77.

17. National Cholesterol Education Program. Report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents. Pediatrics 1992;89(3):525-77.
18. Yu-Poth S, Zhao G, Etherton T, Naglak M, Jonnalagadda S, Kris-Etherton PM. Effects of the National Cholesterol Education Programs Step I and Step II dietary intervention programs on cardiovascular disease risk factors: a meta analysis. Am J Clin Nutr 1999;69:632-46.
19. Lauer RM, Obarzanek E, Hunsberger SA, Van Horn L, Hartmuller VW, Barton BA et al. Efficacy and safety of lowering dietary intake of total fat, saturated fat, and cholesterol in children with elevated LDL cholesterol: the Dietary Intervention Study in Children. Am J Clin Nutr 2000;72(suppl):1332S-42S.
20. Obarzanek E, Kimm SYS, Barton BA, Van Horn L, Kwiterovich PO, Simons-Morton DG et al. Long-term safety and efficacy of a cholesterol-lowering diet in children with elevated low-density lipoprotein cholesterol: seven-year results of the Dietary Intervention Study in Children (DISC). Pediatrics 2001;107(2):256-64.
21. Dietz WH. Childhood weight affects adult morbidity and mortality. J Nutr 1998;128:411S-4S.
22. Newman WP, Freedman DS, Voors AW. Relation of serum lipoprotein levels and systolic blood pressure to early artherosclerosis: The Bogalusa Heart Study. N Eng J Med 1986;314:1250-7.
23. Olmedilla B, Granado F. Growth and macronutrient needs of adolescents. Eur J Clin Nutr 2000;54(suppl 1):S11-15.
24. World Health Organization . Prevention in childhood and youth of adult cardiovascular diseases: Time for action. Geneva: Technical Report series 792; 1990.
25. McPherson RS, Hoelscher DM, Alexander M, Scanlon KS, Serdula MK. Dietary assessment methods among school-aged children: Validity and reliability. Prev Med 2000;31:S11-33.
26. Lee RD, Nieman DC. Nutritional assessment. 3rd ed. Boston: McGraw Hill; 2003.
27. Gibson RS. Principles of nutritional assessment. New York: Oxford University Press; 1990.
28. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Am J Clin Nutr 1997;65(suppl):1100S-7S.
29. Domel SB. Self-reports of diet: how children remember what they have eaten. Am J Clin Nutr 1997;65(suppl):1148S-52S.
30. Rockett HRH, Colditz GA. Assessing diets of children and adolescents. Am J Clin Nutr 1997;65(suppl):1116S-22S.
31. Thompson FE, Moler JE, Freedman LS, Clifford CK, Stables GJ, Willett WC. Register of dietary assessment calibration-validation studies: a status report. Am J Clin Nutr 1997;65(suppl):1142S-7S.

32. Buzzard IM, Sievert YA. Research priorities and recommendations for dietary assessement methodology. Am J Clin Nutr 1994;59(suppl):275S-80S.

33. Little P, Barnett J, Kinmonth A-L., Margetts B, Gabbay J, Thompson R et al. Can dietary assessment in general practice target patients with unhealthy diets? Br J Gen Prac 2000;50:43-5.

34. Gans KM, Ross E, Barner CW, Wylie-Rosett J, McMurray J, Eaton C. REAP and WAVE: new tools to rapidly assess/discuss nutrition with patients. J Nutr 2003;133:556S-62S.

35. Corish CA, Flood P, Kennedy NP. Comparison of nutritional risk screening tools in patients on admission to hospital. J Hum Nutr Dietet 2004;17:133-9.

36. Kondrup J, Allison SO, Elia M, Vellas B, Plauth M. ESPEN guidelines for nutrition screening 2002. Clin Nutr 2003;22(4):415-21.

37. Cleveland LE, Cook DA, Krebs-Smith SM, Friday J. Method for assessing food intakes in terms of servings based on food guidance. Am J Clin Nutr 1997;65(suppl):1254S-63S.

38. Krebs-Smith SM, Cleveland LE, Ballard-Barbash R, Cook DA, Kahle LL. Characterizing food intake patterns of American adults. Am J Clin Nutr 1997;65(suppl):1264S-8S.

39. Hann CS, Rock CL, King I, Drewnowski A. Validation of the Healthy Eating Index with use of plasma biomarkers in a clinical sample of women. Am J Clin Nutr 2001;74:479-86.

40. Newby PK, Hu FB, Rimm EB, Smith-Warner SA, Feskanich D, Sampson L et al. Reproducibility and validity of the Diet Quality Index Revised as assessed by use of a food-frequency questionnaire. Am J Clin Nutr 2003;78:941-9.

41. Kant AK. Indexes of overall diet quality: a review. J Am Diet Assoc 1996;96:785-91.

42. Burema J, Van Staveren WA, Van den Brand PA. Validity and reproducibility. In: Cameron ME, Van Staveren WA, editor(s). Manual on methodology for food consumption studies. Oxford: Oxford University Press; 1988 :171-81.

43. Martin GS, Tapsell LC, Denmeade S, Batterham MJ. Relative validity of a diet history interview in an intervention trial manipulating dietary fat in the management of Type II diabetes mellitus. Prev Med 2003;36:420-28.

44. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453-69.

45. Beaglehole R, Bonita R, Kjellström T. Basic epidemiology. Geneva: World Health Organization; 1993.

46. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. Eur J Clin Nutr 1991;45:569-81.

47. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882-90.
48. Jenner DA, Neylon K, Croft S, Beilin LJ, Vandongen R. A comparison of methods of dietary assessment in Australian children aged 11-12 years. Eur J Clin Nutr 1989;43:663-73.
49. Caan B, Coates A, Schaffer D. Variations in sensitivity, specificity, and predictive value of a dietary fat screener modified from Block et al. J Am Diet Assoc 1995 May;95(5):564-9.
50. Block G. A review of validations of dietary assessment methods. Am J Epidemiol 1982;115(4):492-505.
51. Cavadini C, Siega-Riz AM, Popkin BM. US adolescent food intake trends from 1965 to 1996. Arch Dis Child 2000;83:18-24.
52. Kronsberg SS, Obarzanek E, Affenito SG, Crawford PB, Sabry ZI, Schmidt M et al. Macronutrient intake of black and white adolescent girls over 10 years: the NHLBI Growth and Health Study. J Am Diet Assoc 2003 Jul;103(7):852-60.
53. Kelley C, Krummel D, Gonzales EN, Neal WA, Fitch CW. Dietary intake of children at high risk for cardiovascular disease. J Am Diet Assoc 2004;104(2):222-5.
54. Dwyer JT, Evans M, Stone EJ, Feldman HA, Lytle L, Hoelscher D et al. Adolescents' eating patterns influence their nutrient intakes. J Am Diet Assoc 2001;101(7):798-802.
55. Frank GC. Environmental influences on methods used to collect dietary data from children. Am J Clin Nutr 1994;59(suppl):207S-11S.
56. Wildey MB, Pampalone SZ, Pelletier RL, Zive MM, Elder JP, Sallis JF. Fat and sugar levels are high in snacks purchased from student stores in middle schools. J Am Diet Assoc 2000;100(3):319-22.
57. Nicklas TA, Morales M, Linares A, Yang S-J, Baranowski T, De Moor C et al. Children's meal patterns have changed over a 21-year period: The Bogalusa Heart Study. J Am Diet Assoc 2004;104:753-61.
58. Hackett AF, Jarvis SN, Matthews JNS. A study of the eating habits of 11- and 12-year-old children before and one year after the start of a healthy eating campaign in Northumberland. J Hum Nutr Dietet 1990;3:323-32.
59. Johnson B, Hackett AF. Eating habits of 11-14-year-old schoolchildren living in less affluent areas of Liverpool, UK. J Hum Nutr Dietet 1997;10:135-44.
60. Koletzko B, Dokoupil K, Reitmayr S, Weimert-Harendza B, Keller E. Dietary fat intakes in infants and primary school children in Germany. Am J Clin Nutr 2000;72(suppl):1392S-8S.

61. Labadarios D, Steyn NP, MacIntyre U, Swart R, Gericke GJ, Huskisson J et al. The National Food Consumption Survey (NFCS): Children aged 1-9 years, South Africa, 1999. Stellenbosch, South Africa: The National Food Consumption Survey; 2000.
62. Nel JH, Steyn NP. Report on South African food consumption studies undertaken amongst different population groups (1983-2000): average intakes of foods most commonly consumed. Pretoria: South African Department of Health; 2002.
63. Lubbe AM. A survey of the nutritional status of white school children in Pretoria: description and comparative study of two dietary survey techniques. SA Med J 1968;Jun 22:616-22.
64. Butte NF. Fat intake of children in relation to energy requirements. Am J Clin Nutr 2000;72(suppl):1246S-52S.
65. Anonymous. Guidelines for school health programs to promote lifelong healthy eating. J Sch Health 1997;67(1):9-26.
66. Anderson GH, Zlotkin SH. Developing and implementing food-based dietary guidance for fat in the diets of children. Am J Clin Nutr 2000;72(suppl):1404S-9S.
67. Prentice AM, Paul AA. Fat and energy needs of children in developing countries. Am J Clin Nutr 2000;72(suppl):1253S-65S.
68. Uauy R, Mize CE, Castillo-Duran C. Fat intake during childhood: metabolic responses and effects on growth. Am J Clin Nutr 2000;72(suppl):1354S-60S.
69. Olson RE. Is it wise to restrict fat in the diets of children? J Am Diet Assoc 2000;100(1):28-31.
70. Cheung S-T. Possible dangers in a low fat diet: some evidence reviewed. Nutr Health 2000;14:271-80.
71. Lytle LA. In defense of a low-fat diet for healthy children. J Am Diet Assoc 2000;100(1):39-41.
72. Van Horn L. Primary prevention of cardiovascular disease starts in childhood. J Am Diet Assoc 2000;100(1):41-2.
73. Vorster HH, Nell TA, Kumanyika S, Tee ES. Fats and oils - towards more specific quantitative and qualitative guidelines for South Africa? SA J Clin Nutr 2004;17(2):44-52.
74. Zive MM, Berry CC, Sallis JF, Frank GC, Nader PR. Tracking dietary intake in white and Mexican-American children from age 4 to 12 years. J Am Diet Assoc 2002;102(5):683-9.
75. McGill HC, McMahan CA, Herderick EE, Malcom GT, Tracy RE, Strong JP. Origin of atherosclerosis in childhood and adolescence. Am J Clin Nutr 2000;72(suppl):1307S-15S.
76. Stang J. Assessment of nutritional status and motivation to make behavior changes among adolescents. J Am Diet Assoc 2002;102:S13-21S.

77. Bingham SA. The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. Nutr Abstr Rev 1987;57(10):705-42.

78. Whiting SJ, Shrestha RK. Dietary assessment of elementary school-age children and adolescents. J Can Diet Assoc 1993;54(4):193-6.

79. Baranowski T, Simons-Morton BG. Dietary and physical activity assessment in school-aged children: measurement issues. J School Health 1991;61(5):195-7.

80. Block G, Hartman AM. Issues in reproducibility and validity of dietary studies. Am J Clin Nutr 1989;50:1133-8.

81. Livingstone MBE, Robson PJ. Measurement of dietary intake in children. Proc Nutr Soc 2000;59:279-93.

82. Isaksson B. A critical evaluation of the duplicate-portion technique in dietary surveys. Eur J Clin Nutr 1993;47:457-60.

83. Brady LM, Lesauvage SVM, Saini N, Williams CM, Lovegrove JA. Comparison of dietary fat and fatty acid intake estimated by the duplicate diet collection technique and estimated dietary records. J Hum Nutr Dietet 2003;16:393-401.

84. Van Horn LV, Stumbo P, Moag-Stahlberg A, Obarzanek E, Hartmuller VW, Farris RP et al. The Dietary Intervention Study in Children (DISC): dietary assessment methods for 8- to 10-year-olds. J Am Diet Assoc 1993 Dec;93(12):1396-403.

85. Weber Cullen K, Baranowski T, Baranowski J. Computer software design for children's recording of food intake. J Nutr Educ 1998;30(6):405-9.

86. Van Horn LV, Gernhofer N, Moag-Stahlberg A, Farris R, Hartmuller G, Lasser VI et al. Dietary assessment in children using electronic methods: telephones and tape recorders. J Am Diet Assoc 1990;90(3):412-6.

87. Neighbors-Dembereckyi L. Online diet analysis tools: a functional comparison. J Am Diet Assoc 2002 Dec;102(12):1738-42.

88. Kohlmeier L. Gaps in dietary assessment methodology: meal- vs list-based methods. Am J Clin Nutr 1994;59:175S-9S.

89. Smith AF. Cognitive psychological issues of relevance to the validity of dietary reports. Eur J Clin Nutr 1993;47(suppl):6S-18S.

90. Baranowski T, Domel SB. A cognitive model of children's reporting of food intake. Am J Clin Nutr 1994;59(suppl):212S-7S.

91. Friedenreich CM, Slimani N, Riboli E. Measurement of past diet: review of previous and proposed methods. Epidemiol Rev 1992;14:177-96.

92. Koehler KM, Cunnigham-Sabo L, Lambert LC, McCalman R, Skipper BJ, Davis SM. Assessing food selection in a health promotion program: validation of a brief instrument for American Indian children in the Southwest United States. J Am Diet Assoc 2000;100(2):205-11.
93. Baxter SD, Thompson WO. Accuracy by meal component of fourth-graders' school lunch recalls is less when obtained during a 24-hour recall than as a single meal. Nutr Res 2002;22:679-84.
94. Baxter SD, Thompson WO, Davis HC, Johnson MH. 'How do you remember you ate...?': a Delphi technique study to identify retrieval categories from fourth-grade children. J Am Diet Assoc 1997;97:31-6.
95. Dwyer JT, Coleman KA. Insights into dietary recall from a longitudinal study: accuracy over four decades. Am J Clin Nutr 1997;65(suppl):1153S-8S.
96. Baxter SD, Thompson WO, Davis HC. Prompting methods affect the accuracy of children's school lunch recalls. J Am Diet Assoc 2000;100:911-8.
97. Domel Baxter S, Thompson WO, Smith AF, Litaker MS, Yin Z, Frye FHA et al. Reverse versus forward order reporting and the accuracy of fourth-graders' recalls of school breakfast and school lunch. Prev Med 2003;36:601-14.
98. Drewnowski A. Diet image: a new perspective on the food-frequency questionnnaire. Nutr Rev 2001;59(11):370-4.
99. Story M, Neumark-Sztainer D, French S. Individual and environmental influences on adolescent eating behaviours. J Am Diet Assoc 2002 Mar;102(3):40S-51S.
100.Manios Y, Moschandreas J, Hatzis C, Kafatos A. Health and nutrition education in primary schools of Crete: changes in chronic disease risk factors following a 6-year intervention programme. Br J Nutr 2002;88:315-24.
101.Harrell JS, Gansky SA, McMurray RG, Bangdiwala SI, Frauman AC, Bradley CB. School-based interventions improve heart health in children with multiple cardiovascular disease risk factors. Pediatrics 1998 Aug;102:371-80.
102.Dietz WH, Gortmaker SL. Preventing obesity in children and adolescents. Annu Rev Public Health 2001;22:337-53.
103.American Dietetic Association, Society for Nutrition Education, American School Food Service Association. Position statement - Nutrition Services: An essential component of comprehensive school health programs. J Am Diet Assoc 2003;103:505-14.
104.Florencio CA. Developments and variations in school-based feeding programs around the world. Nutr Today 2001;36(1):29-36.
105.Wechsler H, Devereaux RS, Davis M, Collins J. Using school environment to promote physical activity and healthy eating. Prev Med 2000;31:121S-37S.

106.Kubik MY, Lytle LA, Story M. A practical, theory-based approach to establishing school nutrition advisory councils. J Am Diet Assoc 2001;101(2):223-8.
107.School health policy implementation guidelines 2003 (Department of Health), Pretoria: Department of Health.
108.C2005 Revised national curriculum statement. Grades 4-6 (Schools) 2004 (Department of Education, Gauteng Institute for Educational Development, Gauteng Department of Education), Pretoria: Department of Education.
109.James DCS, Adams TL. Curriculum integration in nutrition and mathematics. J Sch Health 1998;68(1):3-6.
110.Bingham SA, Nelson M, Paul AA, Haraldsdottir J, Loken EB, Van Staveren W. Methods for data collection at an individual level. In: Cameron ME, Van Staveren WA, editor(s). Manual on methodology for food consumption studies. Oxford: Oxford University Press; 1988. :53-106.
111.Kohlmeier L, Bellach B. Exposure assessment error and its handling in nutritional epidemiology. Annu Rev Public Health 1995;16:43-59.
112.Hammond J, Nelson M, Chinn S, Rona RJ. Validation of a food frequency questionnaire for assessing dietary intake in a study of coronary heart disease risk factors in children. Eur J Clin Nutr 1993;47:242-50.
113.Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122:51-65.
114.Burley V, Cade J, Margetts B, Thompson R, Warm D. Consensus document on the development, validation and utilisation of food frequency questionnaires. Leeds, UK: Ministry of Agriculture Fisheries and Food; 2000. Report No.: ANO850.
115.Hankin JH. Development of a diet history questionnaire for studies of older persons. Am J Clin Nutr 1989;50:1121-7.
116.Jain M. Culture-specific food frequency questionnaires: development for use in a cardiovascular study. Can J Diet Prac Res 1999;60:27-36.
117.Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95:781-8.
118.Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB et al. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102(2):212-8,223-5.
119.Bellu R, Ortisi MT, Riva E, Banderali G, Cucco I, Giovannini M. Validity assessment of a food frequency questionnaire for school-age children in Northern Italy. Nutr Res 1995;15(8):1121-28.

120.Little P, Barnett J, Margetts B, Kinmonth A-L., Gabbay J, Thompson R et al. The validity of dietary assessment in general practice. J Epidemiol Community Health 1999;53:165-72.
121.Subar AF, Midthune D, Kulldorf M, Brown CC, Thompson FE, Kipnis V et al. Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. Am J Epidemiol 2000;152:279-86.
122.Beaton GH. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodolgy. Am J Clin Nutr 1994;59(suppl):253S-61S.
123.Hogbin MB, Abbott Hess M. Public confusion over food portions and servings. J Am Diet Assoc 1999;99(10):1209-11.
124.Young LR, Nestle M. Variation in perceptions of a 'medium' food portion: implications for dietary guidance. J Am Diet Assoc 1998;98(4):458-9.
125.Smicklas-Wright H, Mitchell DC, Mickle SJ, Goldman JD, Cook A. Foods commonly eaten in the United States, 1989-1991 and 1994-1996: are portion sizes changing? J Am Diet Assoc 2003;103(1):41-7.
126.Nielsen SJ, Popkin BM. Patterns and trends in food portion sizes, 1977-1998. J Am Med Assoc 2003;289(4):450-453.
127.Peregrin T. A super-sized problem: restaurant chains piling on the food. J Am Diet Assoc 2001;101(6):620.
128.Guthrie HA. Selection and quantification of typical food portions by young adults. J Am Diet Assoc 1984;84(12):1440-4.
129.Cypel YS, Guenther PM, Petot GJ. Validity of portion-size measurement aids: a review. J Am Diet Assoc 1997;97(3):289-92.
130.Byrd-Bredbenner C, Schwartz J. The effect of practical portion size measurement aids on the accuracy of portion size estimates made by young adults. J Hum Nutr Dietet 2004;17:351-7.
131.McGuire B, Chambers E, Godwin S, Brenner S. Size categories most effective for estimating portion size of muffins. J Am Diet Assoc 2001;101(4):470-2.
132.Chambers E, McGuire B, Godwin S, McDowell M, Vecchio F. Quantifying portion sizes for selected snack foods and beverages in 24-hour dietary recalls. Nutr Res 2000;20(3):315-26.
133.Venter CS, MacIntyre UE, Vorster HH. The development and testing of a food portion photograph book for use in an African population. J Hum Nutr Dietet 2000;13:205-18.
134.Bolland JE, Yuhas JA, Bolland TW. Estimation of food portion sizes: effectiveness of training. J Am Diet Assoc 1988;88(7):817-21.
135.Howat PM, Mohan R, Champagne C, Monlezun C, Wozniak P, Bray GA. Validity and reliability of reported dietary intake data. J Am Diet Assoc 1994;94(2):169-73.

136.Brown LB, Hsing-Kuan Oler C. A food display assignment and handling food models improves accuracy of college students' estimates of food portions. J Am Diet Assoc 2000;100(9):1063-5.
137.Goodwin RA, Brulé D, Junkins EA, Dubois S, Beer-Borst S. Development of a food and activity record and a portion-size model booklet for use by 6- to 17-year olds: a review of focus-group testing. J Am Diet Assoc 2001;101(8):926-8.
138.Frobisher C, Maxwell SM. The estimation of food portion sizes: a comparison between using descriptions of portion sizes and a photographic food atlas by children and adults. J Hum Nutr Dietet 2003;16:181-8.
139.Kuehneman T, Stanek K, Eskridge K, Angle C. Comparability of four methods for estimating portion sizes during a food frequency interview with caregivers of young children. J Am Diet Assoc 1994;94(5):548-51.
140.Willett W. Food frequency methods. In: Willett W, editor(s). Nutritional epidemiology. New York: Oxford University Press; 1998:74-100.
141.Kim MK, Choi BY. The influence of portion size data on the agreement of classification of individuals according to nutrient estimates by food frequency questionnaire in a rural area of Korea. Nutr Res 2002;22:271-81.
142.Randall E. Measuring food use in school-aged children. J Sch Health 1991;61(5):201-3.
143.Bellu R, Riva E, Ortisi MT, De Notaris R, Santini I, Giovannini M. Validity of a food frequency questionnaire to estimate mean nutrient intake of Italian school children. Nutr Res 1996;16(2):197-200.
144.Vereecken CA, Maes L. A Belgian study on the reliability and relative validity of the Health Behaviour in School-aged Children food frequency questionnaire. Public Health Nutr 2003;6:581-8.
145.Cavadini C, Decarli B, Dirren H, Cauderay M, Narring F, Michaud P-A. Assessment of adolescent food habits in Switzerland. Appetite 1999;32:97-106.
146.Wilson AMR, Lewis RD. Disagreement of energy and macronutrient intakes estimated from a food frequency questionnaire and a 3-day diet record in girls 4 to 9 years of age. J Am Diet Assoc 2004;104(3):373-8.
147.Dennis LK, Snetselaar LG, Nothwehr FK, Stewart RE. Developing a scoring method for evaluating dietary methodology in reviews of epidemiologic studies. J Am Diet Assoc 2003;103:483-7.
148.Feunekes GIJ, Van Staveren WA, De Vries JHM, Burema J, Hautvast JGAJ. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. Am J Clin Nutr 1993;58:489-96.

149.Willett W, Stampfer M, Chu N-F, Spiegelman D, Holmes M, Rimm E. Assessment of questionnaire validity for measuring total fat intake using plasma lipid levels as critaria. Am J Epidemiol 2001;154(12):1107-12.
150.Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire. Am J Epidemiol 1997;146:856-69.
151.Schaefer EJ, Augustin JL, Schaefer MM, Rasmussen H, Ordovas JM, Dallal GE et al. Lack of efficacy of a food-frequency questionnaire in assessing dietary macronutrient intakes in subjects consuming diets of known composition. Am J Clin Nutr 2000;71:746-51.
152.Vandenlangenberg GM, Mares-Perlman JA, Brady WE, Klein BE, Klein R, Palta M et al. Incorporating fat-modified foods into a food frequency questionnaire improves classification of fat intake. J Am Diet Assoc 1997;97(8):860-6.
153.Broadfield E, McKeever T, Fogarty A, Britton J. Measuring dietary fatty acid intake: validation of a food-frequency questionnaire against 7 d weighed records. Br J Nutr 2003;90:215-20.
154.Nelson M. The validation of dietary assessment. In: Margetts BM, Nelson M, editor(s). Design concepts in nutritional epidemiology. Oxford: Oxford University Press; 1997:241-68.
155.Frank GC. Taking a bite out of eating behaviour: food records and food recalls of children. Jnl Sch Health 1991;61(5):198-200.
156.Wold RS, Lopez ST, Pareo-Tubbeh SL, Baumgartner RN, Romero LJ, Garry PJ et al. Helping elderly participants keep 3-day diet records in the New Mexico Aging Process Study. J Am Diet Assoc 1998;98(3):326-32.
157.Carroll RJ, Pee D, Freedman LS, Brown CC. Statistical design of calibration studies. Am J Clin Nutr 1997;65(suppl):1187S-9S.
158.Simons-Morton BG, Baranowski T. Observation in assessment of children's dietary practices. J Sch Health 1991;61(5):204-7.
159.Livingstone MBE, Black AE. Markers of the validity of reported energy intake. J Nutr 2003;133:895S-920S.
160.Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr 1997;65(suppl):1232S-9S.
161.Van 't Veer P, Kardinaal AFM, Bausch-Goldbohm RA, Kok FJ. Biomarkers for validation. Eur J Clin Nutr 1993;47(suppl 2):58S-63S.
162.Van 't Veer P. Measuring nutritional exposures including biomarkers. Proc Nutr Soc 1994;53:27-35.

163.Joachim G. Supply and demand: a framework for explaining variability in dietary intake and its impact on data. Nutr Health 1997;11:289-99.
164.Joachim G. The influence of time on dietary data: differences in reported summer and winter food consumption. Nutr Health 1997;12:33-43.
165.Potosky AL, Block G, Hartman AM. The apparent validity of diet questionnaires is influenced by number of diet-record days used for comparison. J Am Diet Assoc 1990;90(6):810-3.
166.Volatier JL, Turrini A, Welten D. Some statistical aspects of food intake assessment. Eur J Clin Nutr 2002;56(suppl 2):46S-52S.
167.Basiotis PP, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr 1987;117:1638-41.
168.Hebert JR, Ockene IS, Hurley TG, Luippold R, Well AD, Harmatz MG. Development and testing of a seven-day dietary recall. J Clin Epidemiol 1997;50(8):925-37.
169.Falciglia GA, Troyer AG, Couch SC. Dietary variety increases as a function of time and influences diet quality in children. J Nutr Educ Behav 2004;36:77-83.
170.Nelson M, Bingham SA. Assessment of food consumption and nutrient intake. In: Margetts BM, Nelson M, editor(s). Design concepts in nutritional epidemiology. 2nd ed. Oxford: Oxford University Press; 1997:123-69.
171.Gibson RS. Validity in dietary assessment: a review. J Can Diet Assoc 1990;51(1):275-80.
172.Jula A, Seppänen R, Alanen E. Influence of days of the week on reported food, macronutrient and alcohol intake among an adult population in South Western Finland. Eur J Clin Nutr 1999;52:808-12.
173.Potgieter JF, Fellingham SA. Assessment methods for dietary surveys. S Afr Med J 1967 Sep 16:886-90.
174.Crawford PB, Obarzanek E, Morrison J, Sabry ZI. Comparative advantage of 3-day food records over 24-hour recall and 5-day food frequency validated by observation of 9- and 10-year-old girls. J Am Diet Assoc 1994;94:626-30.
175.Rush D. Nutrition screening in old people: its place in a coherent practice of preventive health care. Annu Rev Nutr 1997;17:101-25.
176.Posthauer ME, Dorse BD, Foiles RA, Escott-Stump S, Lysen L, Balugun L. ADA definitions for nutrition screening and nutrition assessment. J Am Diet Assoc 1994;94:838-9.
177.Taylor Bear M, Bradford Harris A. Pediatric nutrition assessment: identifying children at risk. J Am Diet Assoc 1997;97(suppl 2):107S-15S.

178.Splett P, Myers EF. A proposed model for effective nutrition care. J Am Diet Assoc 2001;101(3):357-63.
179.Byers T, Marshall J, Fiedler R, Zielezny M, Graham S. Assessing nutrient intake with an abbreviated dietary interview. Am J Epidemiol 1985;122:41-50.
180.Warneke CL, Davis M, De Moor C, Baranowski T. A 7-item versus 31-item food frequency questionnaire for measuring fruit, juice, and vegetable intake among a predominantly African-American population. J Am Diet Assoc 2001;101(7):774-9.
181.Byers T, Serdula M, Kuester S, Mendlein J, Ballew C, McPherson RS. Dietary surveillance for states and communities. Am J Clin Nutr 1997;65(suppl):1210S-4S.
182.Wilkin D, Hallam L, Doggett M-A. Measures of need and outcome for primary health care. Oxford: Oxford University Press; 1992.
183.De Vos AS, Strydom H, Fouchè CB, Poggenpoel M, Schurink EW. Research at grass roots - A primer for the caring professions. Pretoria: JL van Schaik; 1998.
184.Keller HH, Hedley MR, Brownlee SW. The development of Seniors in the community: risk evaluation for eating and nutrition (SCREEN). Can J Diet Prac Res 2000;61:67-72.
185.Jones JM. The methodology of nutritional screening and assessment tools. J Hum Nutr Dietet 2002;15:59-71.
186.Ward J, Close J, Little J, Boorman J, Perkins A, Coles SJ et al. Development of a screening tool for assessing risk of undernutrition in patients in the community. J Hum Nutr Dietet 1998;11:323-30.
187.Gregoire MJH, Chow S, Spooner GR, Cave AJ, Klemka PJ. A screening instrument to identify adults with nutritional risks in a family medicine practice. Can J Diet Prac Res 1999;60:205-11.
188.O'Donnell MG, Nelson M, Wise PH, Walker DM. A computerized diet questionnaire for use in diet health education 1. Development and validation. Br J Nutr 1991;66:3-15.
189.Goudge DR, Williams A, Pinnington LL. Development, validity and reliability of the Derby Nutritional Score. J Hum Nutr Dietet 1998;11:411-21.
190.Hiller L, Lowery JC, Davis JA, Shore CJ, Striplin DT. Nutritional status classification in the Department of Veterans Affairs. J Am Diet Assoc 2001;101(7):786-92.
191.Schneider SM, Hebuterne X. Use of nutritional scores to predict clinical outcomes in chronic diseases. Nutr Rev 2000;58(2):31-8.
192.Rocandio Pablo AM, Arroyo Izaga M, Ansotegui Alday L. Assessment of nutritional status on hospital admission: nutritional scores. Eur J Clin Nutr 2003;57:824-31.
193.Laporte M, Villalon L, Payette H. Simple nutrition screening tools for healthcare facilities. Can J Diet Prac Res 2001;62:26-34.

194.Thompson FE, Subar AF, Smith AF, Midthune D, Radimer KL, Kahle LL et al. Fruit and vegetable assessment: performance of 2 new short instruments and a food frequency questionnaire. J Am Diet Assoc 2002;102(12):1764-72.
195.Field AE, Colditz GA, Fox MK, Byers T, Serdula M, Bosch RJ et al. Comparison of 4 questionnaires for assessment of fruit and vegetable intake. Public Health 1998;88:1216-8.
196.Marsh T, Weber Cullen K, Baranowski T. Validation of a fruit, juice and vegetable availability questionnaire. J Nutr Educ Behav 2003;35:100-4.
197.Ling AMC, Horwath C, Parnell W. Validation of a short food frequency questionnaire to assess consumption of cereal foods, fruit and vegetables in Chinese Singaporeans. Eur J Clin Nutr 1998;52:557-64.
198.Golan M, Weizman A. Reliability and validity of the Family Eating and Activity Habits Questionnaire. Eur J Clin Nutr 1998;52:771-7.
199.Van Assema P, Brug J, Ronda G, Steenhuis I. The relative validity of a short Dutch questionnaire as a means to categorize adults and adolescents to total and saturated fat intake. J Hum Nutr Dietet 2001;14:377-90.
200.Yaroch AL, Resnicow K, Khan LK. Validity and reliability of qualitative dietary fat index questionnaires: a review. J Am Diet Assoc 2000;100:240-4.
201.Srinath U, Shacklock F, Scott LW, Jaax S, Kris-Etherton PM. Development of MEDFICTS - a dietary assessment instrument for evaluating fat, saturated fat, and cholesterol intake. J Am Diet Assoc 1993;93(suppl):A105.
202.Srinath U, Shacklock F, Shannon BM, Mitchell DC, Kris-Etherton PM, Scott L et al. MEDFICTS - a dietary assessment instrument for evaluating fat, saturated fat and cholesterol intake. Circulation 1993;88(4):3414.
203.Kris-Etherton P, Eisenstat B, Jaax S, Srinath UMA, Scott L, Rader J et al. Validation for MEDFICTS, a dietary assessment instrument for evaluating adherence to total and saturated fat recommendations of the National Cholesterol Education Program Step 1 and Step 2 diets. J Am Diet Assoc 2001;101:81-6.
204.Schwartz CG, Torisky DM, Gloeckner JW, Marrah GW, Nett M. Adapting MEDFICTS food frequency questionnaire to assess dietary fat intake and educate participants at a cholesterol screening on North Carolina. J Am Diet Assoc 2000;100(suppl)(9):A-40.
205.Taylor AJ, Wong H, Wish K, Carrow J, Bell D, Bindeman J et al. Validation of the MEDFICTS dietary questionnaire: a clinical tool to assess adherence to American Heart Association dietary fat intake guidelines. Nutrition Journal [serial online] 2003;2(4):Available from: http://www.nutritionj.com/content/2/1/4.

206.Block G, Clifford C, Naughton MD, Henderson M, McAdams M. A brief dietary screen for high fat intake. J Nutr Educ 1989;21:199-207.

207.Kemppainen T, Rosendahl A, Nuutinen O, Ebeling T, Pietinen P, Uusitupa M. Validation of a short dietary questionnaire and a qualitative fat index for the assessment of fat intake. Eur J Clin Nutr 1993;47:765-75.

208.Murphy SP, Kaiser LL, Townsend MS, Allen LH. Evaluation of validity of items for a food behavior checklist. J Am Diet Assoc 2001;101(7):751-6,761.

209.Kristal AR, Shattuck AL, Henry HJ. Patterns of dietary behavior associated with selecting diets low in fat: reliability and validity of a behavioral approach to dietary assessment. J Am Diet Assoc 1990;90(2):214-20.

210.Birkett NJ, Boulet J. Validation of a food habits questionnaire: poor performance in male manual laborers. J Am Diet Assoc 1995;95:558-63.

211.Connor SL, Gustafson JR, Sexton G, Becker N, Artaud-Wild S, Connor WE. The Diet Habit Survey: a new method of dietary assessment that relates to plasma cholesterol changes. J Am Diet Assoc 1992;92:41-7.

212.Gans KM, Sundaram SG, McPhillips JB, Hixson ML, Linnan L, Carleton RA. Rate your plate: an eating pattern assessment and educational tool used at cholesterol screening and education programs. J Nutr Educ 1993;25(1):29-36.

213.Ammerman AS, Haines PS, DeVellis RF, Strogatz DS, Keyserling TC, Simpson RJ et al. A brief dietary assessment to guide cholesterol reduction in low-income individuals: design and validation. J Am Diet Assoc 1991;91(11):1385-90.

214.Olendzki B, Hurley TG, Hebert JR, Ellis S, Merriam PA, Luippold R et al. Comparing food intake using the Dietary Risk Assessment with multiple 24-hour dietary recalls and the 7-day dietary recall. J Am Diet Assoc 1999;99(11):1433-9.

215.Capps O, Cleveland L, Park J. Dietary behaviors associated with total fat and saturated fat intake. J Am Diet Assoc 2002;102(4):490-6,501-2.

216.Knapp JA, Hazuda HP, Haffner SM, Young EA, Stern MP. A saturated fat/cholesterol avoidance scale: sex and ethnic differences in a biethnic population. J Am Diet Assoc 1988;88(2):172-7.

217.Retzlaff BM, Dowdy AA, Walden CE, Bovbjerg VE, Knopp RH. The Northwest Lipid Research Clinic fat intake scale: validation and utility. Am J Public Health 1997;87(2):181-5.

218.Kristal AR, Abrams BF, Thornquist MD, Disogra L, Croyle RT, Shattuck AL et al. Development and validation of a food use checklist for evaluation of community nutrition interventions. Am J Public Health 1990;80:1318-22.

219.Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index Revised: a measurement instrument for populations. J Am Diet Assoc 1999;99(6):697-704.
220.Gibson RS. Sources of error and variability in dietary assessment methods: a review. J Can Diet Assoc 1987;48(3):150-5.
221.Joachim G. Are reliability, reproducibility and validity the correct terms to assess the correctness of dietary studies? Nutr Health 2001;15:17-28.
222.Paeratakul S, Popkin BM, Kohlmeier L, Hertz-Picciotto I, Guo X, Edwards LJ. Measurement error in dietary data: implications for the epidemiologic study of the diet-disease relationship. Eur J Clin Nutr 1998;52:772-7.
223.Lissner L, Heitmann BL, Lindroos AK. Measuring intake in free-living human subjects: a question of bias. Proc Nutr Soc 1998;57:333-9.
224.Kristal AR, Andrilla CHA, Koepsell TD, Diehr PH, Cheadle A. Dietary assessment instruments are susceptible to intervention-associated response set bias. J Am Diet Assoc 1998;98(1):40-3.
225.Vuckovich N, Ritenbauch C, Taren DL, Tobar M. A qualitative study of participants' experiences with dietary assessment. J Am Diet Assoc 2000;100(9):1023-8.
226.Berg C, Jonsson I, Conner MT, Lissner L. Sources of bias in a dietary survey of children. Eur J Clin Nutr 1998;52:663-7.
227.Kris-Etherton PM, Etherton TD. The impact of the changing fatty acid profile of fats on diet assessment and health. J Food Comp Analysis 2003;16:373-8.
228.Stumbo PJ. Novel software applications using nutrient databases. J of Food Comp Analysis 2003;16:293-7.
229.Wolmarans P, Laubscher JA. Fatty acid intake and the South African food composition tables. SA J Clin Nutr 2004;17(suppl)(2):S43.
230.Tooze JA, Subar AF, Thompson FE, Troiano R, Schatzkin A, Kipnis V. Psychosocial predictors of energy underreporting in a large double labeled water study. Am J Clin Nutr 2004;79:795-804.
231.Ghiselli EE, Campbell JP, Zedeck S. Measurement theory for the behavioral sciences. San Francisco: WH Freeman & Co; 1981.
232.Mouton J, Marais HC. Basic concepts in the methodology of the social sciences. Pretoria: Human Sciences Research Council; 1988.
233.Miller CK, Achterberg CL. Reliability and validity of a nutrition and food-label knowledge test for women with type 2 diabetes mellitus. J Nutr Educ 1999;31:43-8.
234.Leedy PR, Ormrod JE. Practical research planning and design 7th ed. New Jersey: Merrill Prentice Hall; 2001.

235.McCall R, Cotton E. The validation of a nursing nutritional assessment tool for use on acute elderly wards. J Hum Nutr Dietet 2001;14:137-48.
236.Johnson F, Wardle J, Griffith J. The Adolescent Food Habits Checklist: reliability and validity of a measure of healthy eating behaviour in adolescents. Eur J Clin Nutr 2002;56:644-9.
237.Babbie E, Mouton J. The practice of social research. Cape Town: Oxford University Press; 2001.
238.Becker BG, Indik BP, Beeuwkes AM. Dietary intake methodologies - A review. Bethesda, Maryland: University of Michigan Research Institute; 1960.
239.Wolfe WS, Frongillo EA, Cassano PA. Evaluating brief measures of fruit and vegetable consumption frequency and variety: cognition, interpretation, and other measurement issues. J Am Diet Assoc 2001;101(3):311-8.
240.Editorial. Eight faces of validity. J Am Diet Assoc 2000;100:256.
241.Kohlmeier L. Overview of validity, quality control and measurement error issues in nutritional epidemiology. Eur J Clin Nutr 1993;47(suppl):1S-5S.
242.De Vet HCW, Terwee CB, Bouter LM. Current challenges in clinimetrics. J Clin Epidemiol 2003;56:1137-41.
243.Kubena KS. Accuracy in dietary assessment: On the road to good science. J Am Diet Assoc 2000;100:775-6.
244.Marshall JR, Lanza E, Bloch A, Caan B, Caggiula A, Quandt S et al. Indexes of food and nutrient intakes as predictors of serum concentrations of nutrients: the problem of inadequate discriminant validity. Am J Clin Nutr 1997;65(suppl):1269S-74S.
245.McNulty KY, Adams CH, Anderson JM, Affenito SG. Development and validation of a screening tool to identify eating disorders in female athletes. J Am Diet Assoc 2001;101:886-92.
246.Heller L, Fox S, Hell K, Church JA. Development of an instrument to assess nutritional risk factors for children infected with human immunodeficiency virus. J Am Diet Assoc 2000;100(3):323-9.
247.Thorsdottir I, Gunnarsdottir I, Eriksen B. Screening method evaluated by nutritional status measurements can be used to detect malnourishment in chronic obstructive pulmonary disease. J Am Diet Assoc 2001;101(6):648-54.
248.Sorenson AW, Calkins BM, Connoly MA, Diamond E. Comparison of nutrient intake determined by four dietary intake instruments. J Nutr Educ 1985;17:92-9.
249.Morgan KJ, Johnson SR, Rizek RL, Reese R, Stampley GL. Collection of food intake data: an evaluation of methods. J Am Diet Assoc 1987;87(7):888-96.

250.Baranowski T, Smith M, Baranowski J, Wang DT, Doyle C, Lin LS et al. Low validity of a seven-item fruit and vegetable food frequency questionnaire among third-grade students. J Am Diet Assoc 1997;97(1):66-8.
251.Andersen LF, Bere E, Kolbjornsen N, Klepp K-I. Validity and reproducibility of self-reported intake of fruit and vegetable among 6th graders. Eur J Clin Nutr 2004;58:771-7.
252.Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T et al. The food intake recording software system is valid among fourth-grade children. J Am Diet Assoc 2002;102:380-5.
253.Kumanyika SK, Tell GS, Shemanski L, Martel J, Chinchilli VM. Dietary assessment using a picture-sort approach. Am J Clin Nutr 1997;65(suppl):1123S-9S.
254.Weber JL, Lytle L, Gittelsohn J, Cunningham-Sabo L, Heller K, Anliker J et al. Validity of self-reported dietary intake at school meals by American Indian children: the Pathways Study. J Am Diet Assoc 2004;104(5):746-52.
255.Smith KW, Hoelscher DM, Lytle LA, Dwyer JT, Nicklas TA, Zive MM et al. Reliability and validity of the Child and Adolescent Trial for Cardiovascular Health (CATCH) Food Checklist: a self-report instrument to measure fat and sodium intake by middle school students. J Am Diet Assoc 2001;101(6):635-42,647.
256.Sjöberg A, Hulthén L. Assessment of habitual meal pattern and intake of foods, energy and nutrients in Swedish adolescent girls: comparison of diet history with 7-day record. Eur J Clin Nutr 2004;58:1181-9.
257.Brener ND, Billy JOG, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. J Adolesc Health 2003;33:436-57.
258.Frank GC, Hollatz AT, Webber LS, Berenson GS. Effect of interviewer recording practices on nutrient intake - Bogalusa Heart Study. J Am Diet Assoc 1984;84(12):1432-6,9.
259.Kohlmeier L. Future of dietary exposure asessment. Am J Clin Nutr 1995;61(suppl):702S-9S.
260.Haraldsdóttir J. Minimizing error in the field: quality control in dietary surveys. Eur J Clin Nutr 1993;47(suppl 2):19S-24S.
261.Kassam-Khamis T, Nanchahal K, Mangtani P, dos Santos Silva I, McMichael A, Anderson A. Development of an interview-administered food-frequency questionnaire for use amongst women of South Asian ethnic origin in Britain. J Hum Nutr Dietet 1999;12:7-19.
262.Johnson B, Hackett AF, Bibby A, Cross J. An investigation of the face validity of a food intake questionnaire: lessons for dietary advice. J Hum Nutr Dietet 1999;12:307-16.
263.Gracey M. Historical, cultural, political, and social influence on dietary patterns and nutrition in Australian Aboriginal children. Am J Clin Nutr 2000;72(suppl):1361S-7S.

264.Rohrmann S, Klein G. Validation of a short questionnaire to qualitatively assess the intake of total fat, saturated, monounsaturated, polyunsaturated fatty acids, and cholesterol. J Hum Nutr Dietet 2003;16:111-7.
265.Buzzard IM, Stanton CA, Figueiredo M, Fries EA, Nicholson R, Hogan CJ et al. Development and reproducibility of a brief food frequency questionnaire for assessing fat, fibre, and fruit and vegetable intakes of rural adolescents. J Am Diet Assoc 2001;101(12):1438-46.
266.Cotton PA, Subar AF, Friday JE, Cook A. Dietary sources of nutrients among US adults, 1994 to 1996. J Am Diet Assoc 2004;104(6):921-30.
267.Jonsson I, Gummeson L, Conner M, Svensson E. Assessing food choice in school children: reliability and construct validity of a method stacking food photographs. Appetite 1998;30:25-37.
268.Thompson FE, Byers T. Dietary assessment resources manual. J Nutr 1994;124(suppl)(11S):2245S-2311S.
269.Royal College of Paediatrics and Child health: Ethics Advisory Committee . Guidelines for the ethical conduct of medical research involving children. Arch Dis Child 2000;82:177-82.
270.Altman DG. Practical statistics for medical research. London: Chapman & Hall/CRC; 1991.
271.Willett W. Issues in analysis and presentation of dietary data. In: Willett W, editor(s). Nutritional epidemiology. 2nd ed. Oxford: Oxford University Press; 1998:321-46.
272.Institute of Medicine, Food and Nutrition Board. Dietary Reference Intakes for energy, carbohydrate, fiber, fat, fatty acids, protein, and amino acids (macronutrients). Washington DC: The National Academies Press; 2002.
273.Margetts BM, Vorster HH, Venter CS. Evidence-based nutrition - the impact of information and selection bias on the interpretation of individual studies. SA J Clin Nutr 2003;16(3):79-87.
274.Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. Brit Med J 2000;320:1240-3.
275.Bellach B. Remarks on the use of Pearson's correlation coefficient and other association measures in assessing validity and reliability of dietary assessment methods. Eur J Clin Nutr 1993;47(suppl 2):S42-S45.
276.Andersen LF, Nes M, Lillegaard IT, Sandstad B, Bjorneboe G-EAa, Drevon CA. Evaluation of a quantitative food frequency questionnaire used in a group of Norwegian adolescents. Eur J Clin Nutr 1995;49:543-54.
277.Metcalf PA, Scragg RKR, Sharpe S, Fitzgerald EDH, Schaaf D, Watts C. Short-term repeatability of a food frequency questionnaire in New Zealand children aged 1-14y. Eur J Clin Nutr 2003;57:1498-1503.

278.Yaroch AL, Resnicow K, Davis M, Davis A, Smith M, Khan LK. Developement of a modified picture-sort food frequency administerd to low-income, overweight, African-American adolescents girls. J Am Diet Assoc 2000;100:1050-6.

279.Rockett HRH, Wolf AM, Colditz GA. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. J Am Diet Assoc 1995;95:336-40.

280.Bland JM, Altman DG. Measurement error and correlation coefficients. Br Med J 1996;313:41-2.

281.Burden ST, Bodey S, Bradburn YJ, Murdoch S, Thompson AL, Sim JM et al. Validation of a nutrition screening tool: testing the reliability and validity. J Hum Nutr Dietet 2001;14:269-75.

282.Cullen KW, Baranowski T, Baranowski J, Hebert D, de Moor C. Pilot study of the validity and reliability of brief fruit, juice and vegetable screeners among inner city African-American boys and 17 to 20 year old adults. J Am Coll Nutr 1999;18:442-50.

283.Rockett HRH, Berkey CS, Colditz GA. Evaluation of dietary assessment instruments in adolescents. Curr Opin Clin Nutr Metab Care 2003;6:557-62.

284.Weber Cullen K, Lara KM, De Moor C. Children's dietary fat intake and fat practices vary by meal and day. J Am Diet Assoc 2002;102(12):1773-8.

285.Frank GC, Nicklas TA, Webber LS, Major C, Miller JF, Berenson GS. A food frequency questionnaire for adolescents: Defining eating patterns. J Am Diet Assoc 1992;92:313-8.

286.Fisher JO, Johnson RK, Lindquist C, Birch LL, Goran MI. Influence of body composition on the accuracy of reported energy intake in children. Obes Res 2000 Nov;8(8):597-603.

287.Bandini LG, Must A, Cyr H, Anderson SE, Spadano JL, Dietz WH. Longitudinal changes in the accuracy of reported energy intake in girls 10-15y of age. Am J Clin Nutr 2003;78:480-4.

288.Joachim G. Sources of variability in the reproducubility of food frequency questionnaires. Nutr Health 1998;12:181-8.

289.Hoelscher DM, Day RS, Kelder SH, Ward JL. Reproducibility and validity of the secondary level School-Based Nutrition Monitoring student questionnaire. J Am Diet Assoc 2003;103:186-94.

290.Gould SM, Anderson J. Development and evaluation of a pictorial version of a WIC allowable foods list. J Am Diet Assoc 1999;99(8):978-80.

291.Raat H, Bonsel GJ, Essink-Bot M-L., Landgraf JM, Gemke RJBJ. Reliability and validity of comprehensive health status measures in children: the Child Health Questionnaire in relation to the Health Utilities Index. J Clin Epidemiol 2002;55:67-76.

292.Torun B, Davies PSW, Liningstone MBE, Paolisso M, Sackett R, Spurr GB. Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. Eur J Clin Nutr 1996;50(suppl 1):37S-81S.

293.O'Connor J, Ball EJ, Steinbeck KS, Davies PSW, Wishart C, Gaskin KJ et al. Comparison of total energy expenditure and energy intake in children aged 6-9 y. Am J Clin Nutr 2001;74:643-9.
294.Livingstone MBE, Robson PJ, Black AE, Coward WA, Wallace JMW, McKinley MC et al. An evaluation of the sensitivity and specificity of energy expenditure measured by heart rate and the Goldberg cut-off for energy intake: basal metabolic rate for identifying mis-reporting of energy intake by adults and children: a retrospective anal. Eur J Clin Nutr 2003;57:455-63.
295.Goris AHC, Westerterp-Plantenga MS, Westerterp KR. Undereating and underrecording of habitual food intake in obese men: selective underreporting of fat intake. Am J Clin Nutr 2000;71:130-4.
296.Domel Baxter S, Thompson WO, Litaker MS, Frye FHA, Guinn CH. Low accuracy and low consistency of fourth-grader's school breakfast and school lunch recalls. J Am Diet Assoc 2002;102(3):386-95.
297.Black AE, Cole TJ. Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. J Am Diet Assoc 2001;101:70-80.
298.Black AE. The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. Eur J Clin Nutr 2000;54:395-404.
299.Bandini LG, Cyr H, Must A, Dietz WH. Validity of reported energy intake in preadolescent girls. Am J Clin Nutr 1997;65(suppl):1138S-41S.
300.Bratteby L-E., Sandhagen B, Fan H, Enghardt H, Samuelson G. Total energy expenditure and physical activity as assessed by the doubly labeled water method in Swedish adolescents in whom energy intake was underestimated by 7-d diet records. Am J Clin Nutr 1998;67:905-11.
301.Cook A, Pryer J, Shetty P. The problem of accuracy in dietary surveys. Analysis of the over 65 UK National Diet and Nutrition Survey. J Epidemiol Commun Health 2000;54:611-6.
302.Goris AHC, Meijer EP, Kester A, Westerterp KR. Use of a triaxial accelerometer to validate reported food intakes. Am J Clin Nutr 2001;73:549-53.
303.Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L et al. Statistical approaches for assessing the relative validity of a food freqeuncy questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutr 2003;6(3):313-21.
304.MacIntyre UE, Venter CS, Vorster HH, Steyn HS. A combination of statistical methods for the analysis of the relative validation data of the quantitative food frequency questionnaire used in the THUSA study. Public Health Nutr 2001;4:45-51.
305.Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. J Clin Epidemiol 2000;53:499-503.
306.Matheson DM, Hanson KA, McDonald TE, Robinson TN. Validity of children's food portion estimates: a comparison of 2 measurement aids. Arch Pediatr Adolesc Med 2002;156:867-71.

307.Samet J, Alberg AJ. Surrogate sources of dietary information. In: Willett W, editor(s). Nutritional epidemiology. 2nd ed. Oxford: Oxford University Press; 1998:157-73.
308.Margetts BM, Nelson M. Overview of the principles of nutritional epidemiology. In: Margetts BM, Nelson M, editor(s). Design concepts in nutritional epidemiology. Oxford: Oxford University Press; 1997:3-38.
309.Barratt A, Irwig L, Glasziou P, Cumming RG, Raffle A, Hicks N et al. Users' guides to the medical literature XVII. How to use guidelines and recommendations about screening. J Am Med Assoc 1999;281(21):2029-34.
310.Grimes DA, Schulz KF. Uses and abuses of screening tests. Lancet 2002;359:881-4.
311.Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. Br Med J 2001;323:157-62.
312.Mitchell DC, Smiciklas-Wright H, Friedmann JM, Jensen G. Dietary intake assessed by the Nutrition Screening Initiative Level II Screen is a sensitive but not a specific indicator of nutrition risk in older adults. J Am Diet Assoc 2002;102(6):842-4.
313.Eaden J, Mayberry MK, Sherr A, Mayberry JF. Screening: the legal view. Public Health 2001;115:218-21.
314.Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? Br Med J 1999;319:1562-5.
315.Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley-Interscience Publ; 2000.
316.Prochaska J, Sallis JF, Rupp J. Screening measure for assessing dietary fat intake among adolescents. Prev Med 2001;33:699-706.
317.Kabagambe E, Baylin A, Allan D, Siles X, Spiegelman D, Campos H. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. Am J Epidemiol 2001;154(12):1126-35.
318.Bryan F, Jones JM, Russell L. Reliability and validity of a nutrition screening tool to be used with clients with learning difficulties. J Hum Nutr Dietet 1998;11:41-50.
319.Garrow JS. Validation of methods for estimating habitual diet: proposed guidelines. Eur J Clin Nutr 1995;49:231-2.
320.Simon MS, Lababidi S, Djuric Z, Uhley V, Depper J, Kresge C et al. Comparison of dietary assessment methods in a low-fat dietary intervention program. Nutr Cancer 2001;40(2):108-17.
321.Vorster HH, Oosthuizen W, Jerling JC, Veldman FJ, Burger HM. The nutritional status of South Africans - A review of the literature from 1975-1996. Durban: Health Systems Trust; 1997.

ADDENDA

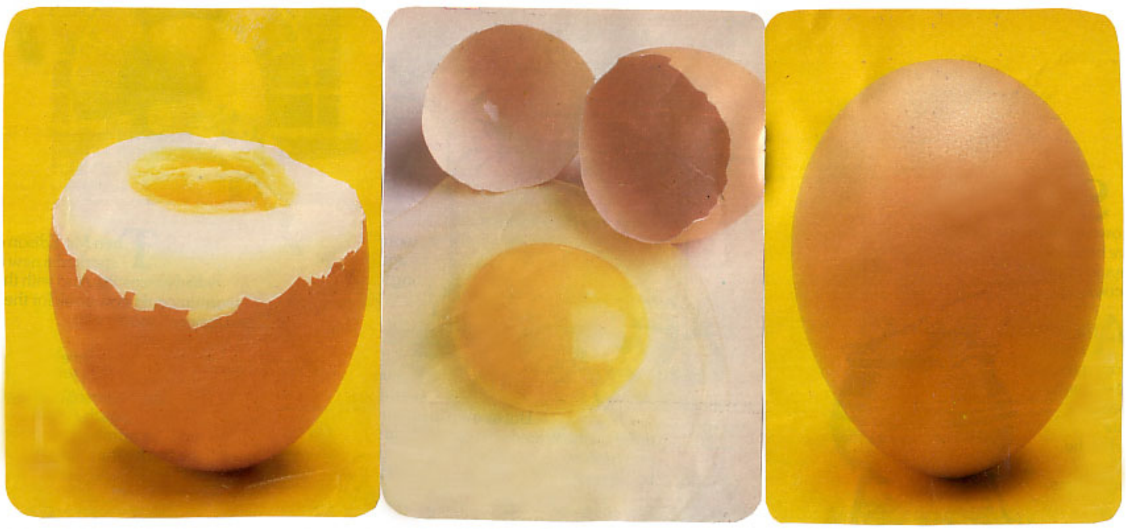
APPENDUM A: TEST METHOD

(Reduced copies of food pictures)

Meats


RAINBOW FARMS
Farm Fresh
FRESH CHICKEN
RENOWN
PORK SAUSAGES
STREAKY BACON
FRANKFURTERS
RED WIENERS
FROZEN CHICKEN LIVERS

Eggs



Dairy, milk, high fat

Dairy, milk, low fat

Dairy, cheese, high fat

Dairy, cheese, low fat

Dairy, dessert, high fat

Dairy, dessert, low fat

Fried foods

Fats In baked goods

Convenience foods

Table fats, high fat

Table fats, low fat

Snacks, high fat

Snacks, low fat

ADDENDUM B: FOOD RECORD FORM

(English translation and Afrikaans version as used in research)

(Instructions and page 1 of day 1; Three pages per day; three days each printed on different colour paper; examples filled in on page 1)

Food recording

Instructions for completion:

General:

1. Each learner keeps record of all food and drink consumed for 3 days
2. Each learner receives a set of record forms (attached) onto which the information must be written. Each day of recording has a different colour paper. Per day there are three pages, thus for day 1 there are 3 pages, and also for days 2 and 3. Make sure you use the right form for each day.
3. Learners in group A will receive an electronic scale for the 3 days.
4. All children who have a kitchen scale at home and have permission to use it, are in group B.
5. Group C consists of children who are neither in group A or group B. They receive a set of measuring cups and spoons and a ruler for the duration of the recording.
6. Following "your" 3 days of recording, you have to hand in your form and the measuring equipment to the mathematic teacher for the next group of learners.
7. When all learners have completed the recording, the collected information will be analysed as a class assignment. It is important that everybody follows the instructions.

Specific instructions:

1. During the recording period you should continue eating as usual. Do not change your eating habits.
2. Everything that crosses your lips (food, sweets, special products, pills, water, cooldrinks) must be written down.
3. In column A you write <u>when</u> and <u>where</u> you have eaten, eg "6:00" "home", "at school", "restaurant".
4. In column B you fill in what you have eaten, eg "sandwich with butter, cheese and tomato", breakfast cereal with milk and sugar".

   Only one food item should be written per line, eg bread, butter, cheese and tomato are each written on a separate line of column B (thus 4 lines).
5. In column C the amounts consumed should be indicated. You should only write down how much you have actually eaten. You should thus measure the food when it is ready to be eaten: eg after wrappers have been removed, after the banana has been peeled. If there are left-overs, eg plate waste or the core of an apple, deduct it. You are welcome to write down your calculation on the form.

<u>Group A:</u> Everything must be recorded to the closest 2gram, exactly as your mathematics teacher showed you. Remember to use the TARE function on the scale.

<u>Group B:</u> Write down the information as precise as your mother's scale allows.

<u>Group C:</u> Try to convert the amount as accurate as possible by holding it next to the household measures you got, ie a cup, spoon, or measure with the ruler.

The following are examples:
Fractions and multiples of a given measure: ½ cup rice
Description: 1 level teaspoon; 2 heaped tablespoons
Dimensions: 2 cookies (each diameter 50mm); 3 rusks (each 100mm x 40mm x 40mm)
Units: 1 big apple; 1 tin of Coke
Packaging: 1 large Bar One, 1 meat pie (place wrappers into the attached plastic pocket; marks it 1, 2 or 3 for the corresponding day)

6. Put all packaging (chips bags, sweets wrappers etc) in the plastic pocket, marked for that applicable day.
7. In column D the food must be described in detail. The attached diagrams will help you remember to note everything precisely

Name and surname:..Grade:.................................... Type of scale:...

Food record form

DAY 1: Day of week:.. Date:...............................

A		B	C	D
How late?	Where?	Whicat food or fluid? (One item per row)	How much did you really eat of each item in B?	Detailed description of each item in B (see flow diagrams)

Naam en van:...Graad:.................................... Tipe skaal:...

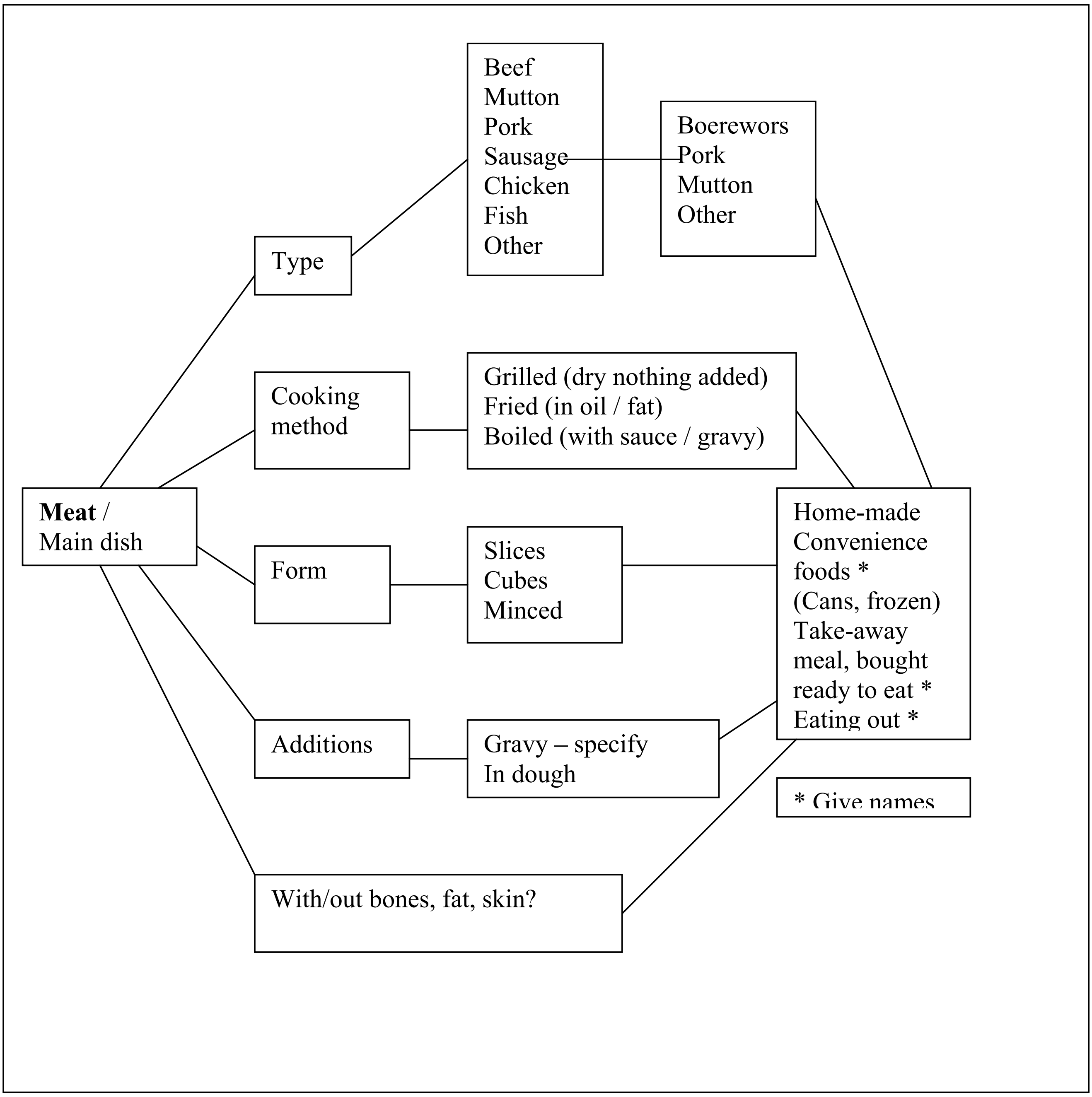
Voedselrekordvorm

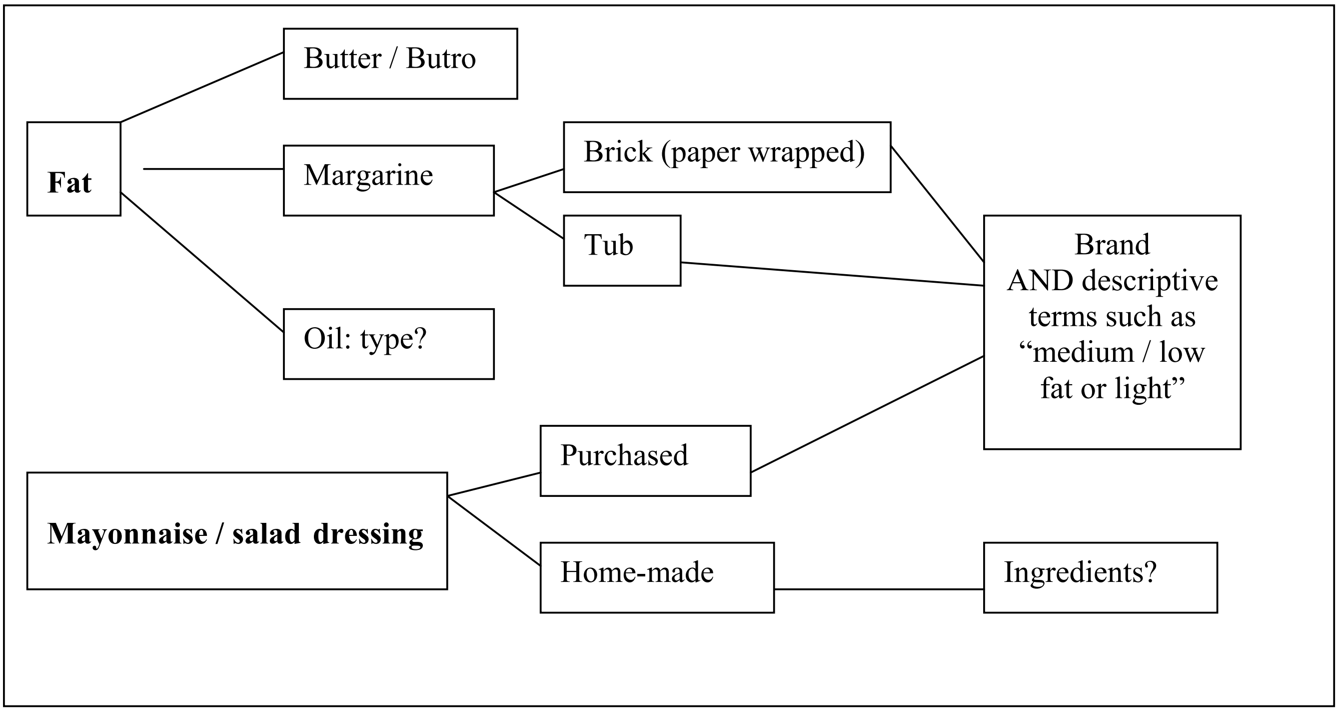
DAG 1: Dag van week:.. Datum:...............................

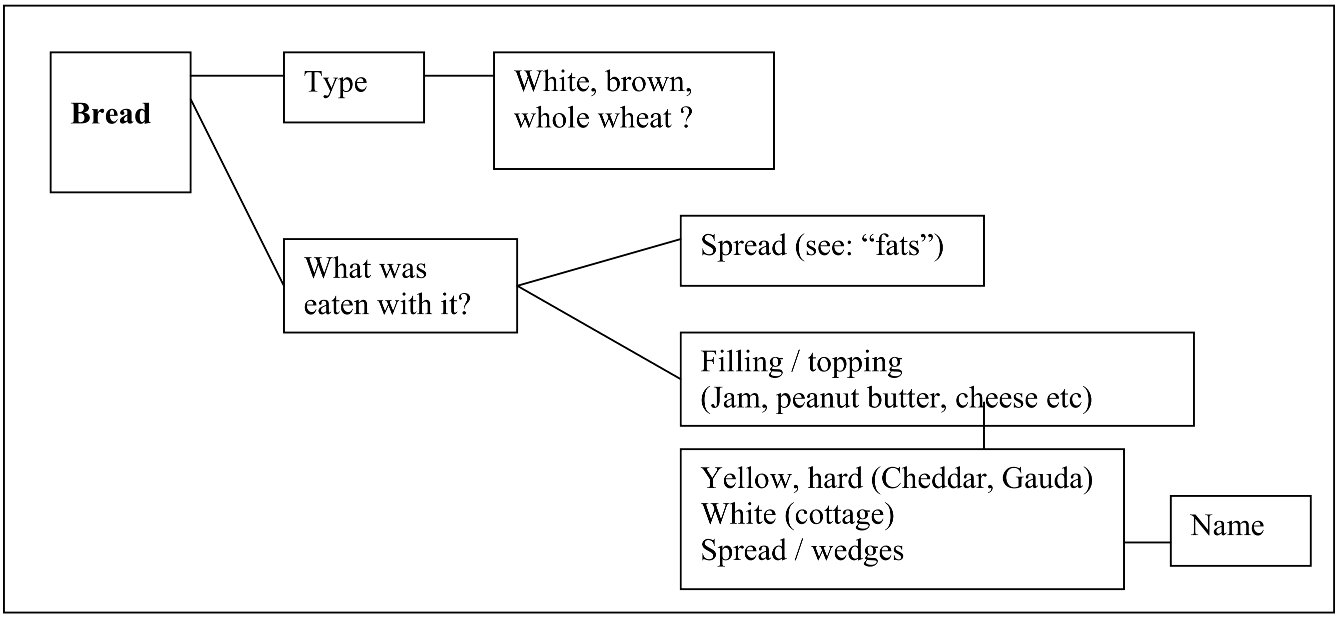
A		B	C	D
Hoe laat?	Waar?	Watter voedsel of vloeistof? (Een item per reël)	Hoeveel het jy regtig geëet van elke item in B?	Volledige beskrywing van elke item in B (kyk vloeidiagramme)

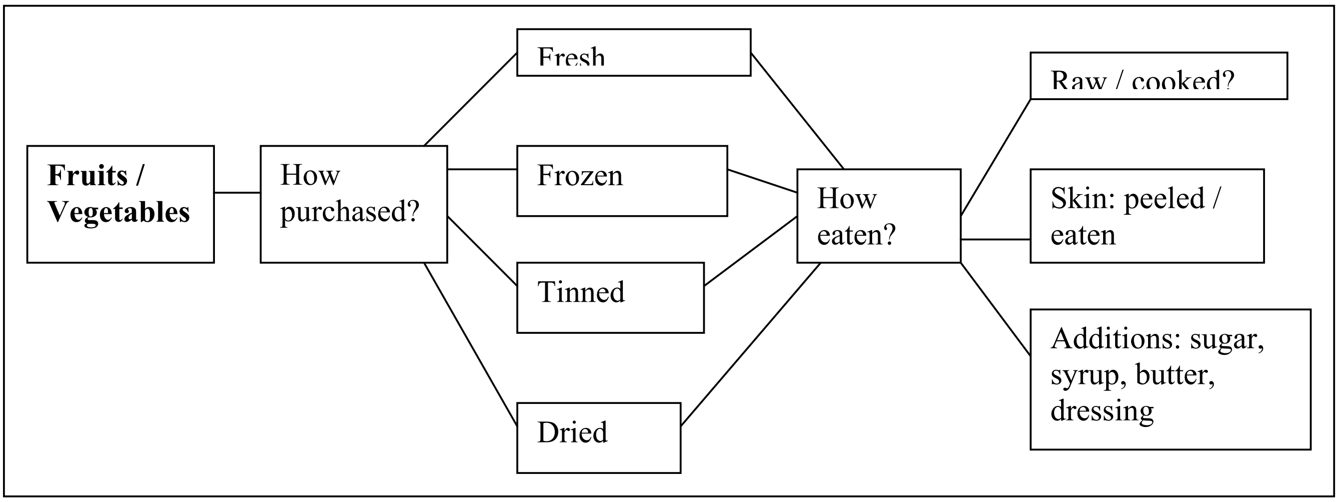
ADDENDUM C: FLOW CHARTS FOR COMPLETING FOOD RECORD

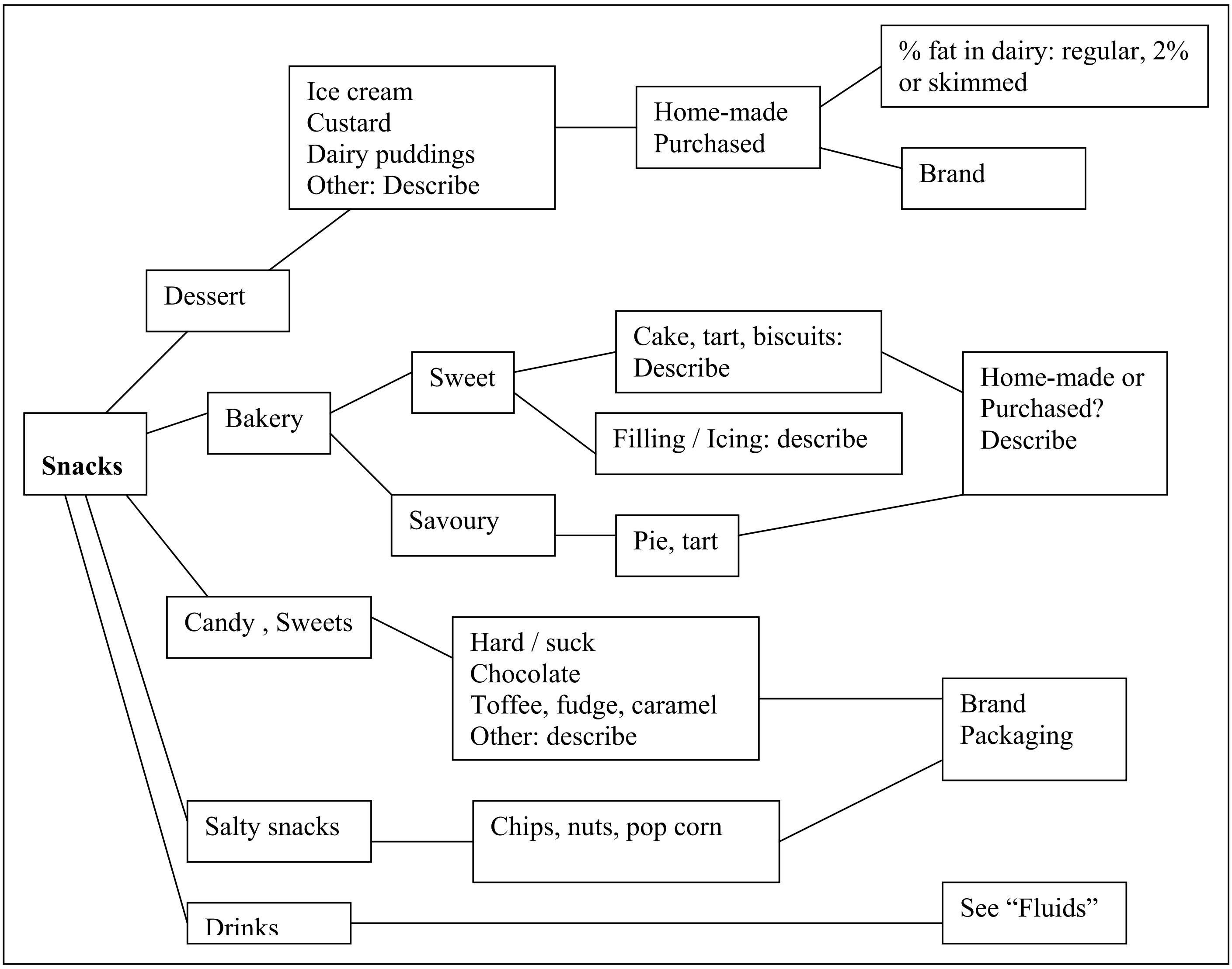
(English translation and Afrikaans version as used in research)

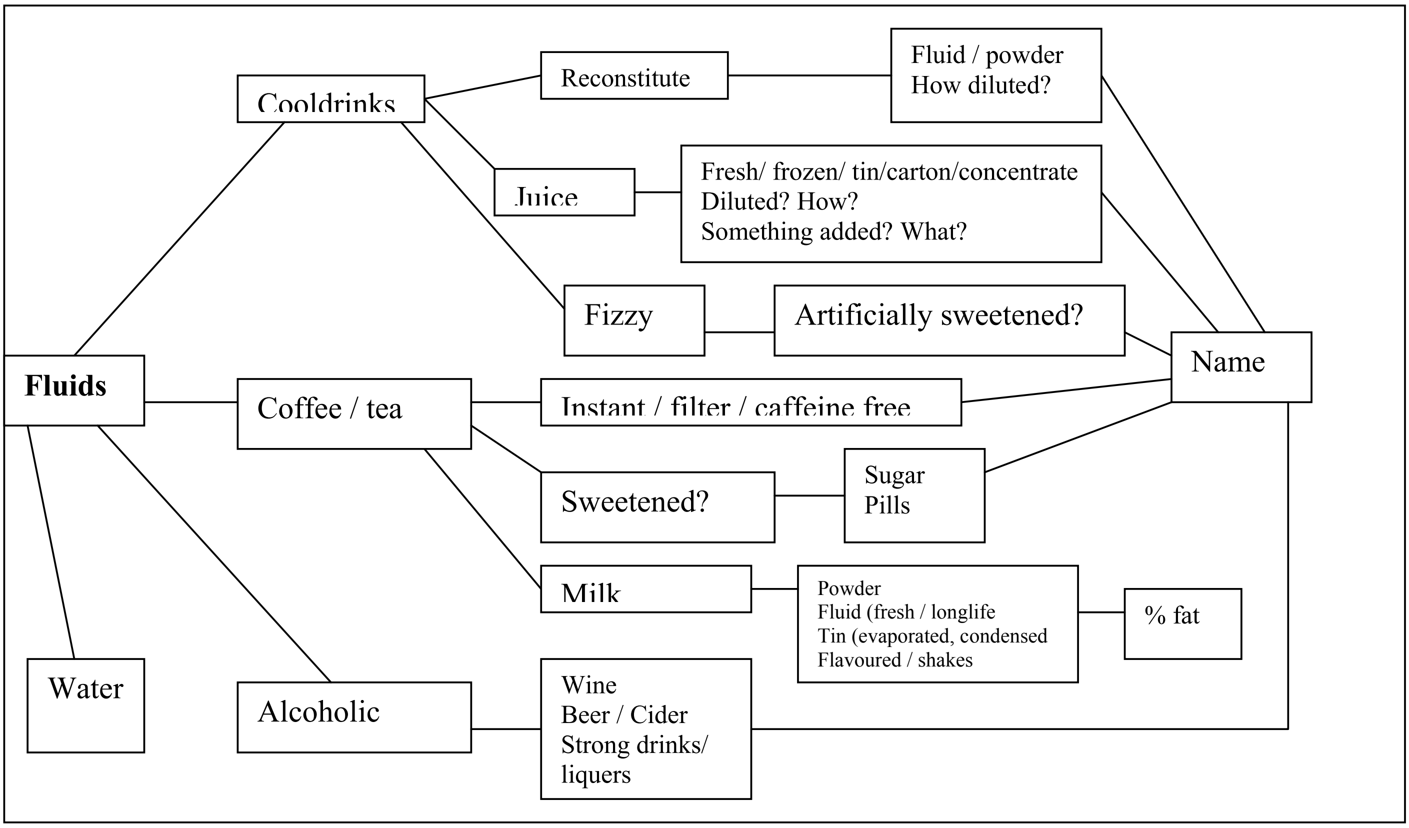

Meat /
Main dish
Type
Beef
Mutton
Pork
Sausage
Chicken
Fish
Other
Boerewors
Pork
Mutton
Other
Cooking method
Grilled (dry nothing added)
Fried (in oil / fat)
Boiled (with sauce / gravy)
Form
Slices
Cubes
Minced
Additions
Gravy – specify
In dough
With/out bones, fat, skin?
Home-made
Convenience foods *
(Cans, frozen)
Take-away meal, bought ready to eat *
Eating out *
* Give names


Fat
Butter / Butro
Margarine
Oil: type?
Brick (paper wrapped)
Tub
Brand AND descriptive terms such as “medium / low fat or light”
Mayonnaise / salad dressing
Purchased
Home-made
Ingredients?


Bread
Type
White, brown, whole wheat ?
What was eaten with it?
Spread (see: “fats”)
Filling / topping (Jam, peanut butter, cheese etc)
Yellow, hard (Cheddar, Gauda)
White (cottage)
Spread / wedges
Name


Fruits / Vegetables
How purchased?
Fresh
Frozen
Tinned
Dried
How eaten?
Raw / cooked?
Skin: peeled / eaten
Additions: sugar, syrup, butter, dressing


Snacks
Dessert
Ice cream
Custard
Dairy puddings
Other: Describe
Home-made
Purchased
% fat in dairy: regular, 2% or skimmed
Brand
Bakery
Sweet
Cake, tart, biscuits: Describe
Filling / Icing: describe
Home-made or Purchased? Describe
Savoury
Pie, tart
Candy , Sweets
Hard / suck
Chocolate
Toffee, fudge, caramel
Other: describe
Brand
Packaging
Salty snacks
Chips, nuts, pop corn
Drinks
See "Fluids"


Fluids
Cooldrinks
Reconstitute
Fluid / powder
How diluted?
Juice
Fresh/ frozen/ tin/carton/concentrate
Diluted? How?
Something added? What?
Fizzy
Artificially sweetened?
Name
Coffee / tea
Instant / filter / caffeine free
Sweetened?
Sugar
Pills
Milk
Powder
Fluid (fresh / longlife
Tin (evaporated, condensed
Flavoured / shakes
% fat
Water
Alcoholic
Wine
Beer / Cider
Strong drinks/
liquers

ADDENDUM D: SCREENER FOR PARENTS

(English translation and Afrikaans version as used in research)

Overview of your grade 6 child's eating habits **Name (child):…………………………….**	**How often?** Fill in the number of times that your child usually eats or drinks this food group per day OR per week Per day	 Per week	**How much?** Mark the typical portion size of your child relative to the given medium (as specified in food list) **S**mall **M**edium **L**arge
Meat (Medium portion: ½ cup, 90g) *Beef, pork, mutton, chicken, sausage, organ meats, processed (cold) meat*			S M L
Eggs			**How many eggs per time?** ≤1 2 ≥3
Milk, full cream (Medium portion: 1 cup) *Full cream, 2%, fresh, longlife, powder, condensed, evaporated; creamers; full cream yoghurt: plain or flavoured*			S M L
Milk, low fat (Medium portion: 1 cup) *Skimmed or 1% melk (fresh, longlife or powder); yoghurt: plain or flavoured*			S M L
Cheese, full cream (Medium portion: Match box size) *Cheddar, Gouda, Tussers, cream cheese, cheese spread and wedges*			S M L
Cheese, low fat (Medium portion: Match box size) *Low fat hard cheeses, low fat cheese spread, wedges and cottage cheese)*			S M L
Dairy dessert, full cream (Medium portion: ½ cup) *Ice cream, milk shakes, custard, blancmange, mousse*			S M L
Dairy dessert, low fat (Medium portion: ½ cup) *Low fat milk drinks, frozen desserts*			S M L
Fried foods (Medium portion: ½ cup) *Fried meat, seafood, fish, chicken; fried vegetables e.g. onions, french fries*			S M L
Bakery (Medium portion: 1 unit) *Cakes, biscuits, sweet and savoury pies, rusks, muffins*			S M L
Convenience foods (Medium portion: 1 cup) *Tins (e.g. meat, soup), ready to eat packages or frozen meals or dishes (e.g. pizza and pasta dishes)*			S M L
Spreads and sauces, regular (Medium portion: 1 teaspoon) *Butter, brick (paper-wrapped) margarine, mayonnaise, regular salad dressing, peanut butter*			S M L
Spreads and sauces, low fat (Medium portion: 1 teaspoon) *Tub margarine (regular, medium or low fat), low fat salad dressing*			S M L
Snacks, high fat (Medium portion: 50g chocolate, 30g [small package] chips, 6 crackers) *Chocolate, peanuts, (crisp) chips, regular salty crackers (eg TUC), caramel/toffee/fudge, coconut*			S M L
Snacks, low fat (Medium portion: 10 sweets, 1 fruit bar, 3 Pro Vitas) *Hard or jelly-type or marshmallow sweets, dried fruit rolls or bars, dry biscuits*			S M L

Kits-oorsig oor u graad 6 kind se eetgewoontes **Naam (kind):…………………………..**	**Hoe dikwels?** Vul die aantal keer in wat u kind die voedsel(groep) gewoonlik per dag OF per week eet/drink Per dag	 Per week	**Hoeveel?** **Merk die tipiese porsiegrootte van u kind relatief tot die gegewe medium (soos aangedui in voedsellys)** **K**lein **M**edium **G**root
Vleis (Mediumporsie: ½ koppie, 90g) *Bees, vark, skaap, hoender, wors, orgaanvleis, geprosesseerde (koue)vleis*			K M G
Eiers			**Hoeveel eiers per keer?** ≤1 2 ≥3
Melk, volroom (Mediumporsie: 1 koppie) *Volroom, 2%, vars, langlewe, poeier, gekondenseerd, ingedamp; verromers; volroomjoghurt; gewoon of gegeur*			K M G
Melk, laevet (Mediumporsie: 1 koppie) *Afgeroomde ("skim") of 1% melk (vars, langlewe of poeier) of joghurt, gewoon of gegeur*			K M G
Kaas, volroom (Mediumporsie: vuurhoutjiedoosgrootte) *Cheddar, Gouda, Tussers, roomkaas, kaassmeer en –wiggies*			K M G
Kaas, laevet (Mediumporsie: vuurhoutjiedoosgrootte) *Laevet harde kaas, laevet kaassmeer, -wiggies en maaskaas*			K M G
Suiwelnagereg, volroom (Mediumporsie: ½ koppie) *Roomys, melkskommels, vla, melkpoedings (blancmange / mousse)*			K M G
Suiwelnagereg, laevet (Mediumporsie: ½ koppie) *Laevet melkdrankies, bevrore nageregte*			K M G
Gebraaide voedsels (Mediumporsie: ½ koppie) *Gebraaide vleis, seekos / vis, hoender; gebraaide groente (bv uie, "slap"-skyfies)*			K M G
Gebak (Mediumporsie: 1 eenheid) *Koekies, koek, soet- en soutterte, pasteie, beskuit, muffins*			K M G
Geriefsvoedsel (Mediumporsie: 1 koppie) *Blikkies (bv vleis, sop), eetgereed pakkies of bevrore maaltye en geregte (bv pizza, pastageregte)*			K M G
Smere en souse, gewoon (Mediumporsie: 1 teelepel) *Botter, blok(papierverpakte) margarien, mayonnaise, gewone slaaisouse, grondboontjiebotter*			K M G
Smere en souse, laevet (Mediumporsie: 1 teelepel) *Bakkiemargarien (gewoon-, medium- of laevet), laevet slaaisouse*			K M G
Versnaperinge, hoëvet (Mediumporsie: 50g sjokolade, 30g [klein pakkie] chips, 6 beskuitjies) *Sjokolade, grondboontjies, (kraak)skyfies, gewone soutbeskuitjies (bv TUC). karamel/toffie/fudge, klapper*			K M G
Versnaperinge, laevet (Mediumporsie: 10 lekkers, 1 vrugtestaaf, 3 Pro Vitas) *Suig-, jellie-tipe-, marshmallow-lekkers, droëvrugterolle of –stawe, droë beskuitjies (broodtipe)*			K M G

ADDENDUM E: INFORMATION LETTER AND INFORMED CONSENT FORM

Ingeligte Toestemming

VOEDINGRISIKOBEPALING BY KINDERS

Navorsingstudie

Ek, ..., gee hiermee toestemming dat my seun/dogter, ..., mag deelneem aan bogenoemde studie soos aan my verduidelik in meegaande dekbrief van mev FAM Wenhold van die Afdeling Mensvoeding, Universiteit van Pretoria.

Doel

Die doel van die studie is om 'n voedingsiftingsvraelys te valideer. Hierdie vraelys is daarop gemik om voedingrisikofaktore by Suid-Afrikaanse laerskoolkinders op 'n geldige, maar koste-effektiewe manier te bepaal.

Beskrywing van prosedures

U moet verstaan dat hierdie studie navorsing behels. U kind sal aan die hand van voedselprente 'n vraelys voltooi. Verder sal u kind 'n drie-dag-rekord hou van alle voedsel wat genuttig word. Laasgenoemde vorm deel van praktiese werk in Wiskunde (uitkoms-gebaseerde onderrig in kurrikulum 2005) en sal ook daar ge-evalueer word. U kind se massa en lengte sal bepaal word. Laastens word u versoek om 'n kitsoorsig oor u kind se eetgewoontes te verskaf.

Risiko en ongemak

Geen

Kontakpersoon vir navrae:
Mev FAM Wenhold (Tel 012-354 1234)
Afdeling Mensvoeding
Fakulteit Gesondheidswetenskappe
Universiteit van Pretoria

Voordele

Om 'n kind se voedingrisiko te ken, maak vroeë intervensie en voorkoming van voedingverwante siektes moontlik. Sou u belangstel, kan die voorlopige resultate ten opsigte van u kind aan u beskikbaar gestel word. Inligting van Stephanus Roos verseker dat stedelike Afrikaanssprekende kinders se eetgewoontes geldig in die projek verteenwoordig word.

Integrasie van weeg- en meetaktiwiteite in die Wiskunde-kurrikulum is in lyn met die filosofie van uitkoms-gebaseerde onderrig en maak leer baie meer sinvol. Verder sal kinders die geleentheid kry om hulle data te verwerk en te interpreter ter bevordering van hulle rekenkundige vaardighede.

Vrywillige deelname

U kind se deelname is vrywillig. Geen vergoeding sal verskaf word nie.. U mag u kind op enige stadium onttrek. Om deelname te weier, sal geen nadelige gevolge vir u kind inhou nie, alhoewel hy/sy die punte verbonde aan die wiskunde-take sal verbeur.

Vertroulikheid

Alle inligting sal vertroulik hanteer word en geen inligting waarvolgens u of u kind geïdentifiseer sou kon word sal vrygestel of gepubliseer word nie.

Ek het al bogenoemde gelees, het die geleentheid gehad om vrae te vra en het bevredigende antwoorde gekry. Ek gee hiermee toestemming dat my kind aan die studie mag deelneem.

....	....
(Ouer/voog handtekening)	(Datum)
....	....
(Kind handtekening)	(Datum)
....	....
(Navorser handtekening)	(Datum)

ADDENDUM F: ANSWER SHEET

(English translation and Afrikaans version as used in research)

ANSWER SHEET

Name: ....Grade:...

CIRCLE YOUR ANSWER.

	Yes 1 / No 2	How often? (Fill in a number for Aper day@ **OR** for Aper week@ . Not for both!)	HOW MUCH? Small 1 / Medium 2 / Large 3
M	1 2 *M1*	per day: *M3D* per week: *M3W*	1 2 3 *M4*
E	1 2 *E1*	per week: *E2*	1 2 3 *E3*
DM1	1 2 *DM1-1*	per day: *DM1-3D* per week: *DM1-3W*	1 2 3 *DM1-4*
DM2	1 2 *DM2-1*	per day: *DM2-3D* per week: *DM2-3W*	1 2 3 *DM2-4*
DC1	1 2 *DC1-1*	per day: *DC1-3D* per week: *DC1-3W*	1 2 3 *DC1-4*
DC2	1 2 *DC2-1*	per day: *DC2-3D* per week: *DC2-3W*	1 2 3 *DC2-4*
DD1	1 2 *DD1-1*	per day: *DD1-3D* per week: *DD1-3W*	1 2 3 *DD1-4*
DD2	1 2 *DD2-1*	per day: *DD2-3D* per week: *DD2-3W*	1 2 3 *DD2-4*
F	1 2 *F-1*	per day: *F-3D* per week: *F-3W*	1 2 3 *F-4*
I	1 2 *I-1*	per day: *I-3D* per week: *I-3W*	1 2 3 *I-4*
C	1 2 *C-1*	per day: *C-3D* per week: *C-3W*	1 2 3 *C-4*
T1	1 2 *T1-1*	per day: *T1-3D*	1 2 3 *T1-4*
T2	1 2 *T2-1*	per day: *T2-3D*	1 2 3 *T2-4*
S1	1 2 *I-1*	per day: *S1-3D* per week: *S1-3W*	1 2 3 *S1-4*
S2	1 2 *S2-1*	per day: *S2-3D* per week: *S2-3W*	1 2 3 *S2-4*

ANTWOORDBLAD

Naam: ....Graad:..

OMKRING JOU ANTWOORD.

	Ja 1	Nee 2	**Hoe dikwels?** (Vul >n syfer by Aper dag@ **OF** by Aper week@ in. Nie by albei nie!)	**HOEVEEL?** Klein 1	Medium 2	Groot 3
M	1	2	per dag: *M3D*	1	2	3
		M1	per week: *M3W*			*M4*
E	1	2 *E1*	per week: *E2*	1	2	3 *E3*
DM1	1	2	per dag: *DM1-3D*	1	2	3
		DM1-1	per week: *DM1-3W*			*DM1-4*
DM2	1	2	per dag: *DM2-3D*	1	2	3
		DM2-1	per week: *DM2-3W*			*DM2-4*
DC1	1	2	per dag: *DC1-3D*	1	2	3
		DC1-1	per week: *DC1-3W*			*DC1-4*
DC2	1	2	per dag: *DC2-3D*	1	2	3
		DC2-1	per week: *DC2-3W*			*DC2-4*
DD1	1	2	per dag: *DD1-3D*	1	2	3
		DD1-1	per week: *DD1-3W*			*DD1-4*
DD2	1	2	per dag: *DD2-3D*	1	2	3
		DD2-1	per week: *DD2-3W*			*DD2-4*
F	1	2	per dag: *F-3D*	1	2	3
		F-1	per week: *F-3W*			*F-4*
I	1	2	per dag: *I-3D*	1	2	3
		I-1	per week: *I-3W*			*I-4*
C	1	2	per dag: *C-3D*	1	2	3
		C-1	per week: *C-3W*			*C-4*
T1	1	2 *T1-1*	per dag: *T1-3D*	1	2	3 *T1-4*
T2	1	2 *T2-1*	per dag: *T2-3D*	1	2	3 *T2-4*
S1	1	2	per dag: *S1-3D*	1	2	3
		1-1	per week: *S1-3W*			*S1-4*
S2	1	2	per dag: *S2-3D*	1	2	3
		S2-1	per week: *S2-3W*			*S2-4*