International Journal of Obesity Supplements (2015) **5,** S22–S28

www.nature.com/ijosup

ORIGINAL ARTICLE

The reliability and validity of a short food frequency questionnaire among 9–11-year olds: a multinational study on three middle-income and high-income countries

T Saloheimo[1], SA González[2], M Erkkola[1], DM Milauskas[3], JD Meisel[4], CM Champagne[3], C Tudor-Locke[3,5], O Sarmiento[2], PT Katzmarzyk[3] and M Fogelholm[1] for the ISCOLE Research Group

OBJECTIVE: The main aim of this study was to assess the reliability and validity of a food frequency questionnaire with 23 food groups (I-FFQ) among a sample of 9–11-year-old children from three different countries that differ on economical development and income distribution, and to assess differences between country sites. Furthermore, we assessed factors associated with I-FFQ's performance.

METHODS: This was an ancillary study of the International Study of Childhood Obesity, Lifestyle and the Environment. Reliability ($n = 321$) and validity ($n = 282$) components of this study had the same participants. Participation rates were 95% and 70%, respectively. Participants completed two I-FFQs with a mean interval of 4.9 weeks to assess reliability. A 3-day pre-coded food diary (PFD) was used as the reference method in the validity analyses. Wilcoxon signed-rank tests, intraclass correlation coefficients and cross-classifications were used to assess the reliability of I-FFQ. Spearman correlation coefficients, percentage difference and cross-classifications were used to assess the validity of I-FFQ. A logistic regression model was used to assess the relation of selected variables with the estimate of validity. Analyses based on information in the PFDs were performed to assess how participants interpreted food groups.

RESULTS: Reliability correlation coefficients ranged from 0.37 to 0.78 and gross misclassification for all food groups was $< 5\%$. Validity correlation coefficients were below 0.5 for 22/23 food groups, and they differed among country sites. For validity, gross misclassification was $< 5\%$ for 22/23 food groups. Over- or underestimation did not appear for 19/23 food groups. Logistic regression showed that country of participation and parental education were associated ($P \leqslant 0.05$) with the validity of I-FFQ. Analyses of children's interpretation of food groups suggested that the meaning of most food groups was understood by the children.

CONCLUSION: I-FFQ is a moderately reliable method and its validity ranged from low to moderate, depending on food group and country site.

International Journal of Obesity Supplements (2015) **5,** S22–S28; doi:10.1038/ijosup.2015.15

INTRODUCTION

Food frequency questionnaires (FFQs) are the most widely used dietary assessment tools in epidemiological studies.[1] FFQs provide crude estimates of usual consumption based on participants self-report of their consumption frequency of pre-listed foods and beverages. A list of foods is provided with predetermined frequency alternatives. FFQs are easy and quick for participants to complete and processing data is inexpensive.[1] FFQs and other self-report methods can be used also among children from the age of ~ 8 years, as their cognitive skills are adequately developed and they are aware of their food intake.[2] Assessment of food intake in children is the focus of the present paper.

The performance (that is, reliability and validity) of a FFQ should be assessed when applying it to a new population.[3] Measuring diet as accurately as possible is crucial for observing true diet–disease associations. Ideally, validation is a part of a process of learning about and developing methods further. Better correlation coefficients between FFQs and reference methods have been reported in studies that ask no portion sizes, measure a short period of time (previous day or week) and for whom the number of questionnaire items is limited to 20–60 items.[4]

Sex, education, weight status,[2] culture[5] and ethnic group[6] are factors which may be related to the validity of FFQs. Also, misreporting based on social desirability has been observed among adults[7] as well as among children.[8,9] The validity of a dietary assessment method may vary even in populations of close geographic proximity.[10] Hence, the validity of dietary assessment instruments is a special challenge in international, multi-country studies with participants from different regions of the world.

A limited number of European epidemiologic studies involving children have conducted validation studies including several countries.[11–13] In some of these studies, results have been presented only as an aggregate of all countries,[12] whereas others have presented analyses stratified by country.[13] A study conducted in Belgium and Italy showed that the validity of a single FFQ was different in the two countries.[13] Efforts have been made to accommodate FFQs to very specific cultural settings.[14] However, to our knowledge, no multinational FFQ validation

[1]Department of Food and Environmental Sciences, University of Helsinki, Helsinki, Finland; [2]School of Medicine, Universidad de los Andes, Bogotá, Colombia; [3]Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, USA; [4]Department of Industrial Engineering, Faculty of Engineering, Universidad de los Andes, CeiBA Complex Systems Research Center, Bogotá, Colombia and [5]Department of Kinesiology, University of Massachusetts Amherst, Amherst, USA. Correspondence: Dr T Saloheimo, Department of Food and Environmental Sciences, University of Helsinki, P.O. Box 66, Helsinki 00014, Finland.
E-mail: taru.saloheimo@helsinki.fi

studies have been conducted, with a study design involving countries from different regions of the world.

The main aim of this study was to assess the reliability and validity of a 23-item FFQ among a sample of 9–11-year-old children in three countries, and to compare these measures between the country sites. In addition, we wanted to gain new insights on the determinants of the FFQ's validity. First, we assessed whether selected background variables were related to the performance of FFQ. Second, we examined how children interpreted the predefined food groups listed in the FFQ.

MATERIALS AND METHODS

Setting, participants and recruitment

This study assessing the reliability and validity of a FFQ was conducted within the context of a larger 12-country study of 9–11-year-old children, the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE).[15] ISCOLE's aim was to determine the relationships between lifestyle behaviors and obesity in a multinational study of children. The eligibility criteria and sampling methods employed in ISCOLE have been reported elsewhere.[15] ISCOLE's sample included children from semiurban and urban areas, and they were recruited from schools, which were stratified by indicators of socioeconomic status.

The present study uses a subsample of ISCOLE with the objective to assess the reliability and validity of the FFQ used in ISCOLE (I-FFQ). A sample of children aged 9–11 years participated in this substudy from three culturally different ISCOLE country sites: USA (Baton Rouge), Colombia (Bogotá) and Finland (Helsinki, Espoo, Vantaa). These countries differed significantly in their Human Development Index and Gini coefficient, with Finland being a high-income country with a low Gini coefficient, the USA a high-income country with a high Gini coefficient and Colombia a middle-income country with a very high Gini coefficient.[16] Our target was to reach a sample size of 100 participants from each country site. A consecutive sampling strategy was used; all available participants of the ISCOLE study were recruited to the validation substudy until the targeted sample size was completed. Data collection for the present study took place between September 2012 and May 2013.

The entire ISCOLE study and the present substudy were approved by the ethical review boards at the participating institutions. The same children participated in the assessment of both reliability and validity in the present study. Ninety-five percent of invited ISCOLE participants completed the reliability component of the present study. Country-specific response rates for the assessment of reliability were 91% (USA), 100% (Colombia) and 96% (Finland). Figure 1 illustrates the participant flow among children participating in the assessment of validity.

The mean age of the 321 participants was 9.9 years (s.d. 0.6), and 53% of the participants were boys. The prevalence of overweight and obesity using the criteria of the World Health Organization[17] was 28%.

Study design and data collection

I-FFQ was completed twice to assess its test–retest reliability. The first I-FFQ was administered as a part of the protocol of the larger ISCOLE study. Participants completed the I-FFQ as a part of a six-page questionnaire in their classrooms. The second FFQ (I-FFQ-2) was administered on average 4.9 (s.d. 1.6) weeks after the first I-FFQ. I-FFQ-2 was administered as a stand-alone questionnaire to the selected subsample participating in the ancillary study in the three countries. Participants were not aware of the reliability aspect of the study, but technicians were not blinded about the purpose of the study.

The validity of I-FFQ (test method) was assessed by comparing it with 3 days of pre-coded food diary (PFD, reference method). The participants began recording PFDs after the first I-FFQ had been administered (within 1 week). The participants were told that the study purpose was to acquire more information about their food habits. They were not told about the validation objective. Information about background variables was obtained from anthropometric measurements for children and from questionnaires completed by parents or guardians.[15]

Food frequency questionnaire (test method). I-FFQ (English, Spanish and Finnish versions available in Supplementary File 1) had similar food groups as the FFQs used in the National Health Behaviour in School-Aged Children Surveys in Belgium and Italy.[13,18] I-FFQ was adapted to different ISCOLE countries so that while the names of the food groups remained the same, examples of individual food items within food groups reflected cultural or regional variation in food consumption patterns. For instance, food group 'fried foods' had 'chicken wings' as one of the example food items in the USA, and 'empanadas' in Colombia.

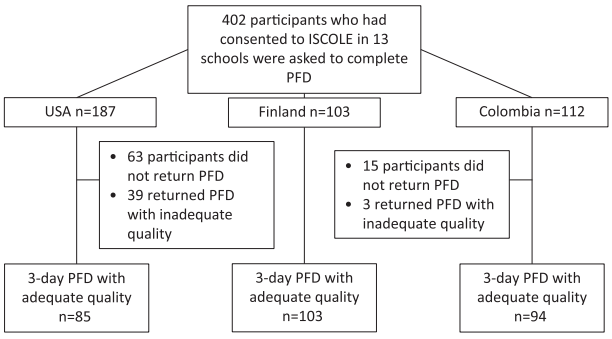

Figure 1. Participant flow in the assessment of validity.

I-FFQ consisted of 23 food and beverage groups. I-FFQ probed for usual consumption with seven frequency alternatives: never; less than once a week; once a week; 2–4 days a week; 5–6 days a week; once a day; and more than once per day. Portion sizes were not collected. I-FFQ was self-administered, but, if needed, participants received assistance from technicians who answered questions in a standardized manner.[15]

Pre-coded food diary (reference method). The PFD served as the criterion in assessing the validity of I-FFQ. A food diary is a suitable method for assessing the validity a FFQ because these methods have the smallest correlated errors.[3] The PFD (Supplementary File 2) used in this study was a form which listed the same 23 food groups as the I-FFQ, with an additional food group called 'other foods'. The PFD was completed prospectively during 3 consecutive days (2 weekdays and a weekend day). The participants used a separate PFD form for each day and they were instructed to fill in the PFD immediately after each meal or snack. They were instructed to enter the consumed food in the most appropriate place in terms of food group and time of the day. The participants were instructed to use the food group 'other foods' if a food item did not fit—according to the participant—in any of the predetermined food groups. Participants were asked to specify the type and brand of consumed food items. Quantities were not asked. The guardians received written instructions on how to assist children in completing the PFDs. Using the PFD, we were able to assess the validity in two ways: first, how well the participants could identify the correct food group, and, second, how well the consumption frequency in the I-FFQ was related to the reported frequency in the PFD.

Data analyses

Validity. In the PFD, the participants recorded the consumed foods in the group which they felt to be correct (PFD reported). Because many of these decisions were incorrect (e.g., diet soda was entered into food group 'regular soda'), the researchers (TS and SG) replaced these food items into the correct food group to compile the PFD corrected. This was subsequently used as the reference in assessing the validity of the frequencies reported in I-FFQ.

The PFD for each day had three time spans (morning, afternoon and evening) and 24 food groups (23 from the I-FFQ and 'other foods'). Having one or more food items recorded into a food group during one time span (that is, one box in the PFD form, see Supplementary File 2) was counted as one eating occasion. Hence, the maximum number of eating occasions (frequency) within any food group was three per day; one in the morning, one in the afternoon and one in the evening. The consumption frequency for the 3 days ranged consequently from 0 to 9.

The I-FFQ data had seven frequency alternatives per week and the PFD-corrected frequencies ranged from 0 to 9 per 3 days. To allow comparison of these two different frequencies, some of the original frequency categories in both data sets were merged (Table 1) to create a similar frequency for both data sets (I-FFQ and PFD corrected): the new frequency categories ranged from 1 to 5 and they were corresponding in terms of weekly and daily consumption frequency. For example, I-FFQ had frequency categories 'once a week' and '2–4 days a week' before variable

Table 1. Variable transformations used for assessing the validity of I-FFQ against PFD corrected

Original variables		*Transformed variables Category used in validity analyses (both in I-FFQ and PFD corrected)*
Consumption frequency in the I-FFQ	*Number of eating occasions on 3-day PFD corrected*[a]	
Never; less than once a week	No eating occasions	1
Once a week; 2–4 days a week	1–3 eating occasions on 1 day	2
5–6 days a week	2–6 eating occasions on 2 days	3
Once a day	3–4 eating occasions on 3 days	4
More than once a day	5–9 eating occasions on 3 days	5

Abbreviations: I-FFQ, food frequency questionnaire used in the International Study of Childhood Obesity, Lifestyle and the Environment; PFD corrected, data set from pre-coded food diary. The consumption of 23 food groups based on researcher's decisions about correct food groups for individual food items ('true food consumption'; see Materials and methods for a detailed explanation). [a]Number of eating occasions on the 3-day PFD ranged from 0 to 9. The daily range was 0–3.

transformations. These two were merged, and the corresponding frequency category in PFD corrected was '1–3 eating occasions on 1 day'. In the PFD, both the number of eating occasions and the number of days, during which the eating occasions appeared, were taken into account. The purpose was to differentiate cases where, for example, 3 eating occasions appeared during 1 day (3 eating occasions per day × 1) from 3 eating occasions appearing on 3 separate days (1 eating occasion per day × 3).

The PFD data were also used to investigate how well the children could place their consumed food items into the food groups listed in the PFD form. In these analyses, the PFD-reported groupings were compared against researchers' decisions about correct food groups. Poor comparability between these two would indicate that the children had problems in interpreting the groupings used in I-FFQ. Investigating decision-making was possible because children had to choose a food group for all consumed food items in the PFD. For example, if a child had consumed an apple cake and written it into food group 'cakes and so on', the decision was considered as correct. If it was written in any other food group, such as 'fruits', the decision was considered as incorrect. Each individual food item was coded to indicate placement in the correct or incorrect category, and the proportion of correct decisions within each food group was calculated. For example, if a child had written 'apple', 'orange' (fruits) and 'carrot' (vegetable) into food group 'fruits', the proportion of correct decisions was 2/3 = 67%. Foods that were correctly reported in category 'other foods' were not counted. If a participant consumed a food item several times a day and always reported it in the same food group, it was still counted as one decision type. The resulting proportion of correct decisions per food group describes all the decisions made during the 3-day PFD.

Statistical methods

Reliability. The reliability of I-FFQ was assessed by comparing it to I-FFQ-2 and by using Wilcoxon signed-rank tests, intraclass correlation coefficients and cross-classification. The existence of systematic error, that is, over- or underestimation (I-FFQ-2 versus initial I-FFQ), was assessed as percentage difference ((I-FFQ − I-FFQ-2)/I-FFQ-2) × 100. The original frequency categories in the I-FFQ were used.

Validity. The agreement between frequencies of I-FFQ and PFD corrected was assessed using Spearman correlation coefficients. The degree of misclassification across five frequency categories between the FFQ and the PFD was examined by cross-classification. The proportion of correctly categorized food groups in the same, adjacent and opposite categories was calculated. Proportion of systematic error was assessed by calculating the percentage difference between test and reference methods ((I-FFQ − PFD corrected)/PFD corrected) × 100.

A logistic regression model was used to investigate the relation of selected variables with an estimate of agreement between I-FFQ and PFD corrected. The selection of characteristics included in the logistic regression model as independent variables was based on evidence from the literature. Body mass index was used as a continuous variable, and parents' highest education as a 3-category variable (low: less than high school or some high school; moderate: completed high school, some college/associate's degree; high: bachelor's degree, postgraduate degree). In the logistic regression analysis, the lowest third of correct recordings (0–8 out of 23) was used as the outcome (the dependent variable).

The validity analyses (Spearman correlation coefficients, cross-classification, percentage difference and logistic regression analysis) were performed using the classification of I-FFQ and PFD-corrected data into five frequency categories as presented in Table 1.

Analyses were completed using SPSS for Windows (SPSS Inc., Chicago, IL, USA, version 22.0) and SAS (SAS Institute Inc., Cary, NC, USA, version 9.3). Figures were presented using Microsoft Powerpoint 2010 (Microsoft Corporation, Redmond, WA, USA) and Adobe Illustrator CC 2014 (Adobe Systems Inc., San Jose, CA, USA).

Children's interpretation of food groups in PFD (PFD reported versus researchers' decisions). For the PFD-reported data, we performed descriptive statistics only, not any inferential analyses. As a result of the data handling, we had the proportion of correct decisions within each food group separately for each participant. To summarize the results, the participants were classified into four groups using the following cutoffs for each food group: 0% correct, 1–49% correct, 50–99% correct and 100% correct. These cutoffs were chosen because we wanted to focus on having none or all correct within each food group, as we considered these decision types as the most significant with regard to the study question. Furthermore, the distribution showed a bimodal distribution, with a clear majority of participants scoring either all or none correct.

RESULTS

Reliability

Results relating to the reliability of I-FFQ are presented in Table 2. According to the Wilcoxon signed-rank test, there was no significant difference between I-FFQ and I-FFQ-2 in 7 out of 23 food groups: 'fruit juice', 'regular soda', 'diet soda', 'low-fat milk', 'whole milk', 'sweets' and 'chips'. Intraclass correlation coefficients ranged from 0.37 ('dark-green vegetables' and 'orange vegetables') to 0.78 ('sports drinks'). Apart from 'orange vegetables', 'sports drinks', 'ice cream', 'meat alternatives', 'fish' and 'deep fried foods', the median consumption frequencies of food groups were within a ± 10% difference between test and retest measurements. In cross-classifications of consumption frequencies by both methods, the greatest misclassification (4.7% in the opposite category) was found for 'whole milk'.

Validity

According to PFD corrected, the proportion of participants who consumed foods from specific groups during the 3-day period of PFD varied from 2.8% (energy drinks) to 84.0% (fruits; Supplementary Table 1).

Validity of reported frequencies. Figure 2 displays Spearman correlation coefficients between I-FFQ and PFD corrected, which varied between participating countries and food groups (Supplementary Table 2). Correlation coefficients for the three countries combined ranged from 0 ('fruit juice') to 0.64 ('low-fat milk').

Table 2. Test–retest reliability of I-FFQ: consumption frequency[a] in I-FFQ compared with ISCOLE-FFQ-2 ($n = 321$)

	Median (25th and 75th percentile)		P-*value*	*ICC*	*% Difference*	*% (n) Gross misclassification*
	I-FFQ	*I-FFQ-2*				
Fruits	5 (4;6)	4 (4;6)	0.043	0.48	−3.7	0.3 (1)
Vegetables	4 (3;6)	4 (3;5)	< 0.001	0.52	−6.9	0.9 (3)
Dark-green vegetables	3 (2;4)	3 (2;4)	0.019	0.37	−7.6	2.2 (7)
Orange vegetables	4 (2;4)	3 (2;4)	< 0.001	0.37	−11.6	1.3 (4)
Fruit juice	5 (3;7)	5 (3;7)	0.137	0.64	−2.5	0.6 (2)
Regular soda	3 (2;4)	3 (2;4)	0.504	0.61	−1.5	0.0 (0)
Diet soda	1 (1;2)	1 (1;2)	0.133	0.40	−5.3	1.6 (5)
Energy drinks	1 (1;1)	1 (1;1)	0.036	0.68	−8.5	0.6 (2)
Sports drinks	2 (1;4)	1 (1;3)	< 0.001	0.78	−15.2	0.0 (0)
Low-fat milk	3 (1;6)	3 (1;7)	0.487	0.72	2.8	2.5 (8)
Whole milk	4 (1;6)	3 (1;6)	0.613	0.57	−1.4	4.7 (15)
Cheese	4 (2;5)	4 (2;5)	0.002	0.50	−7.6	1.3 (4)
Ice cream	3 (2;4)	3 (2;4)	< 0.001	0.53	−12.5	0.9 (3)
Other milk products	4 (4;6)	4 (3;5)	< 0.001	0.40	−9.5	0.6 (2)
Sweets	4 (3;5)	3 (3;5)	0.363	0.57	−1.8	0.0 (0)
Cakes and so on	2 (2;3)	2 (2;3)	0.027	0.47	−5.8	0.0 (0)
Chips	3 (2;4)	3 (2;4)	0.223	0.47	−3.5	0.6 (2)
French fries	3 (2;4)	2 (2;3)	0.002	0.51	−7.3	0.0 (0)
Whole-grain bread and so on	4 (3;6)	4 (2;6)	0.006	0.50	−6.5	0.6 (2)
Meat alternatives	4 (3;5)	4 (2;5)	< 0.001	0.53	−10.9	0.9 (3)
Fish	3 (2;4)	2 (2;3)	< 0.001	0.49	−11.4	0.6 (2)
Deep fried foods	3 (2;4)	3 (2;4)	< 0.001	0.47	−14.1	0.6 (2)
Fast foods	3 (2;4)	3 (2;3)	< 0.001	0.63	−9.5	0.3 (1)

Abbreviations: I-FFQ, food frequency questionnaire used in the International Study of Childhood Obesity, Lifestyle and the Environment; ICC, intraclass correlation coefficient; I-FFQ-2, ISCOLE-FFQ-2 was administered after the initial I-FFQ to assess the test–retest reliability of the method; *P*, Wilcoxon *P*-value for the comparison of I-FFQ and I-FFQ-2; % difference, (I-FFQ − I-FFQ-2)/I-FFQ-2) ×100. [a]Consumption frequency: 1, never; 2, less than once a week; 3, once a week; 4, 2–4 days a week; 5, 5–6 days a week; 6, once a day, every day; and 7, every day, more than once.

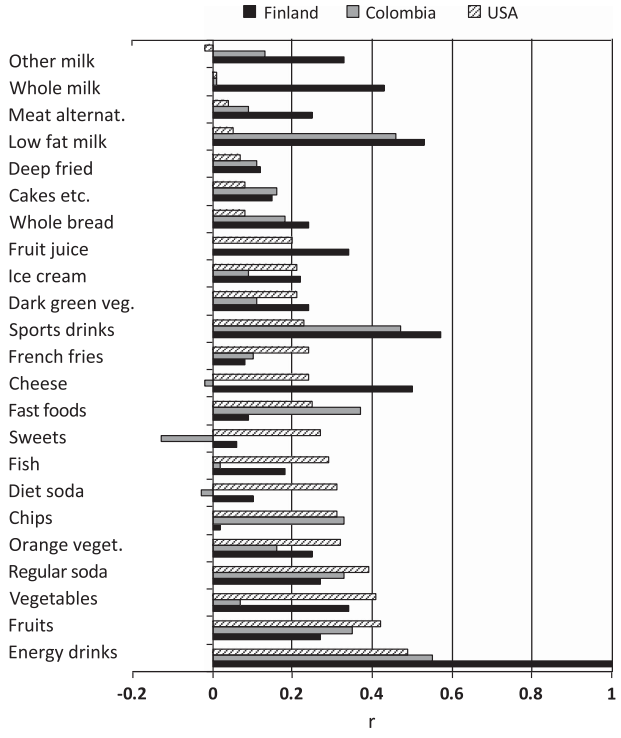

Figure 2. Validity of I-FFQ against PFD corrected: correlation coefficients ($n = 282$).

Figure 3 illustrates the percentage of subjects categorized into the categories 'same', 'adjacent', 'opposite' and 'other' by the I-FFQ and PFD (Supplementary Table 3). Food groups 'energy drinks', 'diet soda' and 'sports drinks' had the highest proportion of participants in category 'same'. 'Fruit juice', 'fruits' and 'other milk products' had the lowest proportion of participants in the category 'same'. 'Fruit juice' had the largest proportion of participants in category 'opposite'.

The percentage difference was calculated to assess whether participants had systematically over- or underestimated their food consumption frequency in I-FFQ compared with PFD corrected. The median difference was 0% for 19 out of 23 food groups for the three participating countries combined. Overestimation (median %) was observed for 'dark-green vegetables' (50%), 'orange vegetables' (13%) and 'fruit juice' (100%). Underestimation appeared for 'cakes and so on' (−33%). The trend of over/underestimation was similar across countries. The tendency to over-report was highest among participants in the USA.

A logistic regression ($n = 257$) analysis was performed to assess the association between participant's body mass index and sex, parents' highest education and country of participation on the likelihood that a participant belonged to the lowest third of correct recordings (all food groups together; min–max 0–23 correct recordings) in I-FFQ as compared with PFD corrected (Table 3). Hence, this analysis was used to search for correlates of poor validity on an individual level.

Colombian and Finnish children had lower odds (79% and 93%, respectively) of having the least amount of correct recordings in I-FFQ compared with children in USA. Children of moderately educated parents (that is, those who completed high school or some college/associate's degree) and those of parents with the least education (that is, those who completed less than high school or some high school) were more likely to have incorrect recordings compared with participants with a highly educated parent (that is, those with bachelor's degree or postgraduate degree). BMI was not related to the validity.

Children's interpretation of food groups in PFD (PFD reported versus researchers' decisions). Taking all the food groups together, the

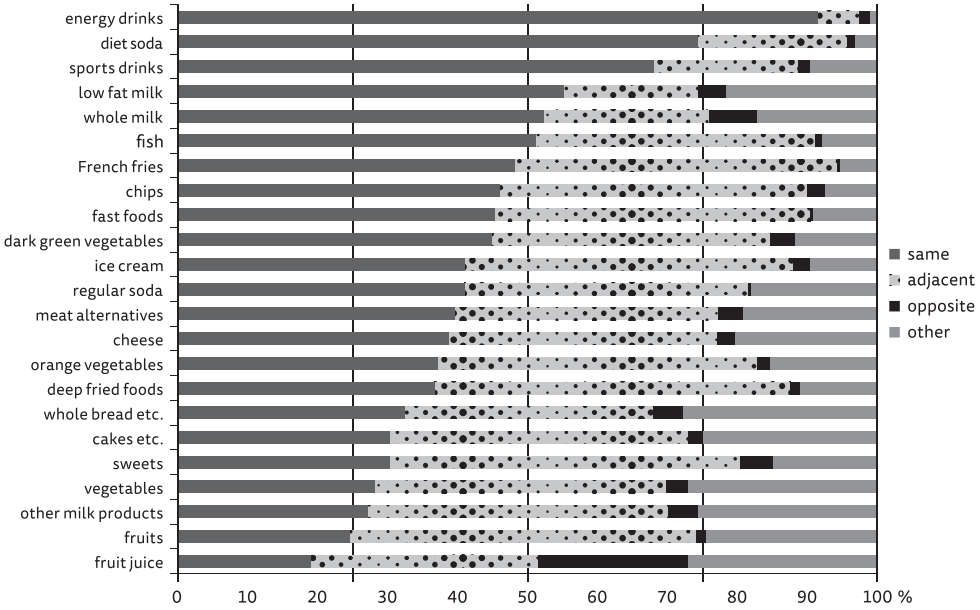

Figure 3. Validity of I-FFQ against PFD corrected: cross-tabulations ($n = 282$).

Table 3. Adjusted odds ratios and 95% confidence intervals for the lowest third (0–8) of correct recordings in I-FFQ ($n = 257$).

	OR	*95% CI*
Sex		
Boy	1.00 (ref.)	
Girl	1.31	0.74, 2.30
BMI	0.99	0.91, 1.10
Highest parental education		
High	1.00 (ref.)	
Moderate	2.62	1.11, 6.22
Low	4.21	1.38, 12.86
Country		
USA	1.00 (ref.)	
Colombia	0.21	0.114, 0.73
Finland	0.07	0.03, 0.17

Abbreviations: CI, confidence interval; I-FFQ, food frequency questionnaire used in the International Study of Childhood Obesity, Lifestyle and the Environment; low amount of correct readings in I-FFQ, 0–8 correct recordings out of 23 in I-FFQ when compared with the reference method (pre-coded food diary); OR, odds ratio; parental education: low: less than high school or some high school; moderate: completed high school, some college/associate's degree; high: bachelor's degree, postgraduate degree; ref., reference category. The model includes all covariates presented in the table. The goodness of fit (Nagelkerke R Square): 0.257.

mean (s.d.) proportion of decision types classified as incorrect was 41.9 (9.5)%, 40.5 (15.0)% and 23.3 (11.4)% in USA, Colombia and Finland, respectively. Figure 4 illustrates that participants understood the meaning of food groups defined by researchers to varying degrees (Supplementary Table 4). For 19 out of 23 food groups, more than half of participants had placed all foods correctly. Diet soda, dark-green vegetables and fruit juice were the most difficult food groups because more than half of participants had no correct reporting in these food groups. Also, 'whole-grain bread and so on' was poorly understood: only 37% had all recorded in the correct food group and 33% had no correct reporting at all. Among the most frequently consumed foods, 'fruits' were the best interpreted food group: 84% reported all of their fruit consumption correctly and only 2% of participants had no correct recording within this food group.

DISCUSSION

The reliability and validity of the ISCOLE food frequency questionnaire (I-FFQ) for 9–11-year-old children varied by food group and by country site. For reliability, intraclass correlation coefficients were 0.37 or above. For validity, Spearman correlation coefficients were below 0.5 for most food groups relative to a criterion standard of PFD. However, percentage of gross misclassification relative to the criterion standard was below 5% for all food groups with a single exception ('fruit juice'). Systematic over- or underestimation was not seen for most food groups. Parents' higher education was associated with better performance in I-FFQ. Children interpreted most food groups in PFD correctly relative to researchers' choices.

There is no FFQ exactly identical to the I-FFQ whose performance would have been assessed among a similar population. Two studies are the most relevant comparison points for our reliability and validity analyses, both by Vereecken *et al.* from the Health Behaviour in School-aged Children survey.[13,18]

In their first study, Vereecken *et al.*[18] assessed the reliability and validity of a 15-item FFQ among 11–12-year-old Belgian children. This 15-item FFQ had 11 food or beverage groups that were very similar to the food groups used in the I-FFQ. For the reliability of the 15-item FFQ, correlation coefficients in the 11 relevant food groups ranged from 0.44 (other milk products) to 0.70 (semi-skimmed milk). For the validity of the 15-item FFQ, correlation coefficients for the relevant 11 food groups ranged from 0.10 (crisps) to 0.65 (semi-skimmed milk). The FFQ overestimated the consumption of most food groups relevant to our study (11 out of 15 food groups), the largest overestimation being for diet soft drinks (658%). The percentage of gross misclassification (once a week or less versus 5 or more times per week) ranged from 1% (chips) to 21% (diet soft drinks).

In a more recent study from Vereecken *et al.*,[13] the validity and reliability of a 14-item FFQ with the same 11 comparable food groups with I-FFQ was assessed in a sample of Belgian and Italian children aged 11–12 years. Reliability correlation coefficients of the 14-item FFQ for the 11 relevant food groups ranged from 0.38 (other milk products in Belgium) to 0.83 (vegetables in Italy).

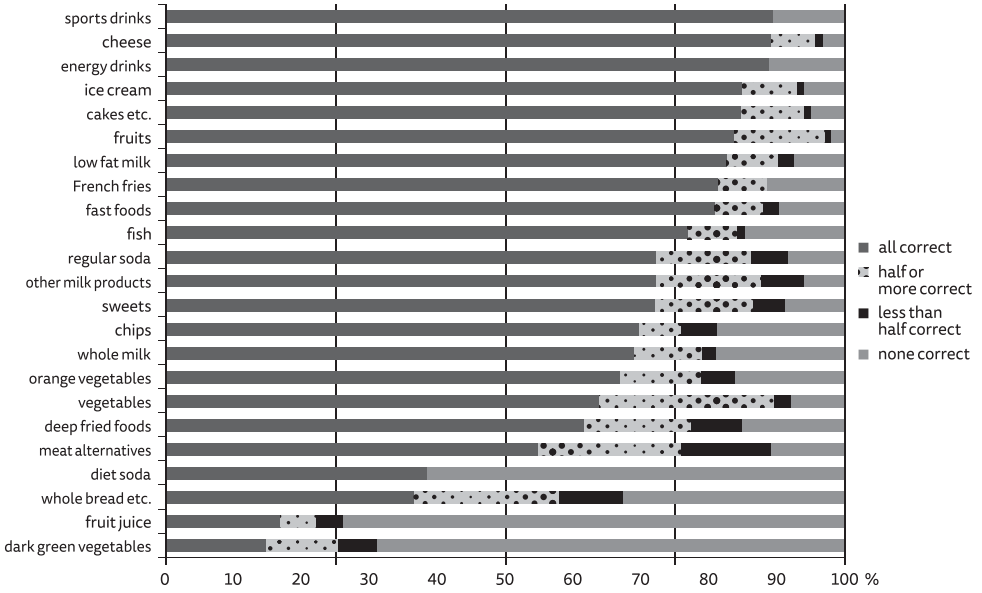

Figure 4. Participants' interpretation of food groups in PFD reported ($n = 282$).

The lowest validity correlation coefficient for the 11 relevant food groups in the 14-item FFQ was reported for diet soft drinks in Italy (−0.13), and the highest for semi-skimmed milk in Italy (0.59). Over-reporting by the 14-item FFQ occurred for most of the 11 food groups.

In the present study, the reliability and validity correlation coefficients of the I-FFQ are comparable to the two studies above. Some validity correlation coefficients were higher in the two Vereecken's studies than in ours.[13,18] The percentage of gross misclassification in I-FFQ was smaller than in Vereecken's first study[18] and the I-FFQ demonstrated less systematic over-reporting than was seen in Vereecken's second study.[13]

Moore *et al.*[8] observed misreporting of dietary intake among children based on social desirability.[8] In our study, the consumption frequency of three food groups representing recommended, healthy foods were slightly overestimated, whereas the food group 'cakes and so on' was underestimated. Body mass index was not associated with agreement between I-FFQ and the reference method. Previous studies suggest that weight status is associated with misreporting more strongly in older children and adolescents compared with younger children.[2]

As in the study by Vereecken *et al.*,[13] estimates of reliability and validity differed between countries in our study. Differences between countries may be due to small differences in administrating the methods, or to true regional and cultural differences. Relevant cultural factors may include children's participation in cooking, knowledge about food and nutrients, and curricula at schools. It is very unlikely that the observed differences would be due to differences in cognitive abilities.

Analyses of children's interpretation of predetermined food groups indicated that participants had no difficulties in understanding what most food groups meant. Yet the validity of consumption frequency of many of the food groups was only poor or moderate. This suggests that the key issue in the validity of I-FFQ may be participants' ability to conceptualize time and consumption frequency. Choosing a frequency category out of 7 weekly alternatives is a difficult cognitive task. Livingstone and Robson[19] pointed out that a limited concept of time is one of the issues that may complicate the dietary assessment of children. As an alternative to the traditional format of a FFQ, future studies could consider using a method in which children are asked to specify the consumption frequency themselves. Another possibility for the low-to-moderate correlations could be that the 3-day period for PFD was too short to capture the normal day-to-day variation of food consumption. However, when designing the study we wanted to choose a method that would not be too burdensome for our young participants.

In our study, especially beverages typically sold by the bottle or can (energy drinks, sports drinks and regular soda) were well interpreted; their validity correlation coefficients were the highest, with the smallest difference between the country sites. It is possible that the consumption pattern of such products is easier for children to conceptualize compared with many other food groups. Food groups that were not well interpreted included assumptions about nutritive value or ingredients ('meat alternatives'/protein content, 'diet soda'/no sugar, 'whole-grain bread and so on'/high fiber content, 'fruit juice'/only fruit as ingredient, 'dark-green vegetables'/not any green vegetable but specific and nutritive dark-green vegetables). These food groups had also large variation in validity correlation coefficients between countries.

Our results suggest that food groups in child FFQs should be formulated based on attributes familiar to children. Criteria related to topics unfamiliar to children should be avoided (cooking method, ingredients and nutritive value). Cooking habits and food knowledge, for example, are also most likely to vary between different regions of the world, whereas some food groups, such as cheese, are quite universal.

The limitation of short FFQs is that they do not cover all foods in an individual's diet. On the other hand, they are designed to capture usual consumption patterns unlike, for example, a 24-h food recall, which measures the food intake of a single day. Besides individual food groups, FFQs can be used for compiling diet scores and indices as has been done in the ISCOLE study.[20]

To conclude, the reliability of the I-FFQ was moderate. Its validity was acceptable for assessing the consumption of most food groups among the 9–11-year-old participants in our international sample. However, the validity of assessment of 'fruit juice' consumption was poor. Our study involved three countries with clearly different dietary habits. The results did not indicate any substantial country site-specific bias in the reliability or validity of the I-FFQ. Moreover, our three-country approach is likely to give a robust projection about the performance of the items in I-FFQ in other similar settings.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the ISCOLE External Advisory Board and the ISCOLE participants and their families who made this study possible. JDM has received funding from the 'Programa nacional de formación doctoral Francisco Jose de Caldas' from Colciencias' (Convocatorias 511-2010). Carlos Mario Pedraza, David Ricardo Leal and Luis Guillermo Gomez are acknowledged for their contribution to this study. A membership list of the ISCOLE Research Group and External Advisory Board is included in Katzmarzyk *et al.* (this Issue). ISCOLE was funded by The Coca-Cola Company. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1 Willett W. Food Frequency Methods. In: Willet W (ed.). *Nutritional Epidemiology,* 3rd edn. Oxford University Press: New York, NY, USA, 2013, pp 70–95.
2 Livingstone MBE, Robson PJ, Wallace JMW. Issues in dietary intake assessment of children and adolescents. *Br J Nutr* 2004; **92** (Suppl 2): S213–S222.
3 Willett W, Lenart E. Reproducibility and validity of food-frequency questionnaires. In: Willet W (ed.). *Nutritional Epidemiology,* 3rd edn. Oxford University Press: New York, NY, USA, 2013, pp 96–141.
4 Kolodziejczyk JK, Merchant G, Norman GJ. Reliability and validity of child/adolescent food frequency questionnaires that assess foods and/or food groups. *J Pediatr Gastroenterol Nutr* 2012; **55**: 4–13.
5 Cassidy CM. Walk a mile in my shoes: culturally sensitive food-habit research. *Am J Clin Nutr* 1994; **59** (Suppl 1): S190–S197.
6 Jaceldo-Siegl K, Fan J, Sabaté J, Knutsen SF, Haddad E, Beeson WL *et al.* Race-specific validation of food intake obtained from a comprehensive FFQ: the Adventist Health Study-2. *Public Health Nutr* 2011; **14**: 1988–1997.
7 Maurer J, Taren DL, Teixeira PJ, Thomson CA, Lohman TG, Going SB *et al.* The psychosocial and behavioral characteristics related to energy misreporting. *Nutr Rev* 2006; **64**: 53–66.
8 Moore GF, Tapper K, Moore L, Murphy S. Cognitive, behavioral, and social factors are associated with bias in dietary questionnaire self-reports by schoolchildren aged 9 to 11 years. *J Am Diet Assoc* 2008; **108**: 1865–1873.
9 Klesges LM, Baranowski T, Beech B, Cullen K, Murray DM, Rochon J *et al.* Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10-year-old African-American girls: results from the Girls Health Enrichment Multisite Studies (GEMS). *Prev Med* 2004; **38**(Suppl 1): S78–S87.
10 Ferguson EL, Gibson RS, Opare-Obisaw C. The relative validity of the repeated 24h recall for estimating energy and selected nutrient intakes of rural Ghanaian children. *Eur J Clin Nutr* 1994; **48**: 241–252.
11 Mouratidou T, Miguel ML, Androutsos O, Manios Y, De Bourdeaudhuij I, Cardon G *et al.* Tools, harmonization and standardization procedures of the impact and outcome evaluation indices obtained during a kindergarten-based, family-involved intervention to prevent obesity in early childhood: the ToyBox-study. *Obes Rev* 2014; **15** (Suppl 3): S53–S60.
12 Bel-Serrat S, Mouratidou T, Pala V, Huybrechts I, Börnhorst C, Fernández-Alvira JM *et al.* Relative validity of the Children's Eating Habits Questionnaire-food frequency section among young European children: the IDEFICS study. *Public Health Nutr* 2013; **17**: 266–276.
13 Vereecken CA, Rossi S, Giacchi MV, Maes L. Comparison of a short food-frequency questionnaire and derived indices with a seven-day diet record in Belgian and Italian children. *Int J Public Health* 2008; **53**: 297–305.
14 Sharma S.. Development and use of FFQ among adults in diverse settings across the globe. *Proc Nutr Soc* 2011; **70**: 232–251.
15 Katzmarzyk PT, Barreira TV, Broyles ST, Champagne CM, Chaput JP, Fogelholm M *et al.* The International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE): design and methods. *BMC Public Health* 2013; **30**: 900.
16 World Bank: World Bank Gini Index. Available at http://data.worldbank.org/indicator/SI.POV.GINI/. 2012.
17 De Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007; **85**: 660–667.
18 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**: 581–588.
19 Livingstone MB, Robson PJ. Measurement of dietary intake in children. *Proc Nutr Soc* 2000; **59**: 279–93.
20 Mikkilä V, Vepsäläinen H, Saloheimo T, Gonzalez SA, Meisel JD, Hu G *et al.* An international comparison of dietary patterns in 9—11-y-old children: the 12-country ISCOLE-study. *Int J Obes Suppl* 2015 (this issue).

Supplementary Information accompanies this paper on *International Journal of Obesity Supplements* website (http://www.nature.com/ijosup)