

# Test–retest reliability and construct validity of the DOiT (Dutch Obesity Intervention in Teenagers) questionnaire: measuring energy balance-related behaviours in Dutch adolescents

Evelien HC Janssen<sup>1</sup>, Amika S Singh<sup>1,\*</sup>, Femke van Nassau<sup>1</sup>, Johannes Brug<sup>2</sup>, Willem van Mechelen<sup>1</sup> and Mai JM Chinapaw<sup>1</sup>

<sup>1</sup>Department of Public and Occupational Health, VU University Medical Center, EMGO Institute for Health and Care Research, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands; <sup>2</sup>Department of Epidemiology and Biostatistics, VU University Medical Center, EMGO Institute for Health and Care Research, Amsterdam, The Netherlands

Submitted 22 December 2011: Final revision received 29 October 2012: Accepted 7 November 2012: First published online 10 December 2012

## Abstract

**Objective:** Adequate assessment of energy balance-related behaviours in adolescents is essential to develop and evaluate effective obesity prevention programmes. The present study examined the test–retest reliability and construct validity of a questionnaire assessing energy balance-related behaviours in adolescents during the evaluation of the DOiT (Dutch Obesity Intervention in Teenagers) intervention.

**Design:** To assess test–retest reliability, adolescents filled in the questionnaire twice ( $n$  111). To assess construct validity, the results from the first test were compared with data collected in a personal cognitive interview ( $n$  20, independent from the reliability study). For both reliability and validity, intraclass correlation coefficients for continuous data or Cohen's kappa coefficients for categorical data were calculated as well as percentage agreement.

**Setting:** Data were collected during school time from February to May 2010.

**Subjects:** Study participants were Dutch adolescents aged 12–14 years attending pre-vocational secondary schools.

**Results:** In more than three-quarters of the ninety-five questionnaire items the test–retest reliability appeared to be good to excellent. Moderate reliability was found for all other twenty-one items. Fifty-one items (of ninety-five items) showed good to excellent construct validity. Construct validity appeared moderate in twenty-three items and poor in twenty-one items. Most items with poor construct validity concerned consumption of sugar-containing beverages and high-energy snacks/sweets.

**Conclusions:** Our study showed good test–retest reliability and largely moderate to good construct validity for the majority of items of the DOiT questionnaire. Items with poor construct validity (most of them found for items concerning energy intake-related behaviours) should be revised and tested again to improve the questionnaire for future use.

**Keywords**  
Reliability  
Validity  
Children  
Diet  
Physical (in)activity

Overweight and obesity among children and adolescents are a major threat to population health<sup>(1)</sup>. Over the past decades unhealthy changes in lifestyles have accelerated due to economic development, industrialisation and globalisation. Dietary patterns are shifting, while energy expenditure is declining<sup>(2)</sup>. Unhealthy diets and lack of physical activity are among the leading causes of the major non-communicable diseases, including CVD, type 2 diabetes and certain types of cancer, and contribute substantially to the global burden of disease, disability and mortality<sup>(3)</sup>.

School-based interventions aimed at promoting healthy diets and physical activity may contribute to prevention of overweight in children and adolescents<sup>(4)</sup>. In 2002 the Dutch Obesity Intervention in Teenagers (DOiT) was developed with the aim of preventing excessive weight gain in Dutch adolescents aged 12–14 years. DOiT especially targets adolescents from lower socio-economic backgrounds. The goal of this comprehensive school-based prevention programme is to induce behavioural changes concerning energy intake and energy expenditure. DOiT focuses on a range of energy balance-related

\*Corresponding author: Email a.singh@vumc.nl

behaviours (EBRB), i.e. consumption of sugar-containing beverages, consumption of high-energy (high-caloric) snacks and sweets, levels of sedentary behaviour and levels of physical activity<sup>(5)</sup>. DOiT proved to be effective in inducing favourable changes in body composition as indicated by sums of skinfold thickness in girls, as well as consumption of sugar-containing beverages in both boys and girls as indicated by self-reports<sup>(6)</sup>.

To measure changes in EBRB and assess the effectiveness of interventions, reliable and valid instruments are vital<sup>(7,8)</sup>. Questionnaires are widely used for the assessment of behavioural changes in population-based studies, since they are inexpensive and easy to administer<sup>(9)</sup>. Unfortunately, most questionnaires focus on one specific behaviour and have limited or unknown psychometric characteristics<sup>(10–12)</sup>. Since there were no reliable and valid questionnaires addressing the whole range of EBRB targeted by the DOiT intervention, we developed the DOiT questionnaire.

Based on its favourable effects in the initial trial<sup>(6)</sup>, we made the DOiT intervention available for all pre-vocational schools in the Netherlands, accompanied by research on implementation and process data<sup>(13)</sup>. The current study aimed to assess the test–retest reliability and construct validity of the DOiT questionnaire.

## Methods

### *DOiT questionnaire*

The self-administered DOiT questionnaire was developed to assess EBRB in adolescents (aged 12–14 years) of pre-vocational secondary schools. The questionnaire was divided into nine sections, i.e. (A) demographic characteristics, (B) consumption of sugar-containing beverages (soft drinks and fruit juices; diet sodas excluded), (C) consumption of high-energy snacks and sweets (high in sugar and fat), (D) breakfast behaviour and frequency of meals, (E) screen behaviour, (F) active transport, (G) physical activity during leisure time, (H) physical activity at school and (I) having a job.

At the beginning of each section, written information was provided on the items of the following section, including examples of soft drinks and fruit juices, for example. This information was also provided orally during the completion of the questionnaire, according to a standardised protocol.

In the current study the test–retest reliability and construct validity of all sections related to diet and physical activity (B to I) were assessed. In total, ninety-five multiple-choice question items were assessed.

Items from the questionnaire used in the first randomised controlled trial evaluating the effectiveness of the DOiT intervention<sup>(6)</sup> (based on other questionnaires<sup>(14–17)</sup>), in combination with items derived from questionnaires of the ENERGY (European Energy balance Research to prevent

excessive weight Gain among Youth) project<sup>(18–20)</sup> and items advised by the Dutch National Institute for Public Health and the Environment<sup>(21)</sup>, were adapted to the target behaviours and population and resulted in the current version of the DOiT questionnaire.

The DOiT questionnaire was pre-tested for comprehensibility and duration of completion among six adolescents from one school, not participating in the present study. After completion of the questionnaire by the adolescents, a structured focus group interview was conducted. Based on the results of this pre-test, no adjustments were required to the questionnaire. The instruction protocol of the questionnaire was adapted in order to clarify the additional information given by the research assistant during completion of the questionnaire.

### *Study population*

For the current study we recruited adolescents aged 12–14 years old, i.e. the target population of the DOiT intervention<sup>(5)</sup>. Pre-vocational secondary schools located in different areas of the Netherlands were recruited to form a representative sample with regard to degree of urbanisation and socio-economic status. Recruitment and data collection took place from February to May 2010. Schools were recruited by email, telephone and a teacher forum on a website. When a school was interested in participation, extra information was provided and dates for measurement were planned. Each participating school selected one class of twenty to twenty-five adolescents for participation. Only adolescents capable of completing a questionnaire in the Dutch language were included in the study. Data collected for the test–retest reliability and construct validity studies were from different adolescents (attending the same school classes). An information letter was sent to the parents of the adolescents. This letter contained a passive informed consent form, meaning that the parents were offered the opportunity to decline participation without signing and returning a consent form. Adolescents and/or their parents who did not want (their child) to participate were excluded from the study. The Medical Ethics Committee of the VU University Medical Center approved the study protocol.

### *Study design and data collection*

#### *Test–retest reliability*

Reliability concerns consistency and reproducibility of measurements. This means that a questionnaire is reliable when measurements done under equal circumstances repeatedly give the same results<sup>(8)</sup>. The reliability of the DOiT questionnaire was examined using a test–retest design. Adolescents were asked to complete the paper-and-pencil self-administered questionnaire in the classroom under the supervision of a trained research assistant. The research assistant guided the class through each section of the questionnaire using a structured protocol.

All concepts were explained in writing in the questionnaire and verbally by the research assistant. In addition, the research assistant brought products to the classroom to serve as examples. The baseline measurements were carried out on all school days except for Mondays, due to the fact that some questions refer to 'yesterday' as a school day. Exactly one week later, on the same weekday and under comparable circumstances, the adolescents were asked to fill in the questionnaire for a second time. The research assistant provided information at the beginning of each section, guiding the adolescents through the questionnaire. Completion of the questionnaire took 45 min on average.

#### *Construct validity*

Construct validity is the extent to which a test measures constructs that it intends to measure and is assessed by comparing the scores of the questionnaire with scores of established measures<sup>(22)</sup>. Because of the absence of a 'gold standard', the construct validity of the DOiT questionnaire was determined by the agreement between the answers of the self-administered questionnaire (first measurement) and a questionnaire completed by a research assistant based on information collected in a personal cognitive interview. This construct validation method has been used previously to validate questionnaires on children's EBRB<sup>(19,23)</sup> and we gain information on whether the adolescents interpreted the questions as intended. For reasons of feasibility, the sample size of the construct validity study was determined to be at least twenty adolescents. The research team asked the teacher of each class participating in the test-retest study to select three to four adolescents, representative for the class, for participation in the construct validity study. Data from these adolescents were excluded from the test-retest reliability study. The participating adolescents were asked to volunteer for a cognitive interview about the same topics as the questionnaire after filling in the first questionnaire. These adolescents filled in the questionnaire together with the other adolescents in the classroom (first measurement) and were subsequently interviewed by a research assistant. The interviews were performed using a standard question route (interview guide), considering the course of an adolescent's day. The interviews took 25 min on average and were audio-recorded and transcribed. Based on the transcribed interview, a second research assistant filled in a second identical questionnaire (second measurement). Both the interviewer and the second research assistant were blinded to the answers of the first questionnaire.

#### *Data management*

All data were entered in the statistical software package SPSS version 18.0 according to a standardised protocol. For both the test-retest reliability and construct validity studies, a randomly selected 5% of the questionnaires were re-entered in SPSS to check for typing errors and

misinterpretation. In cases where there was a difference of more than 3% between the entries, the questionnaires had to be re-entered in the original data set and the procedure was repeated. The rate of disagreement in both studies ranged from 0.0 to 2.0%. Subsequently, data were cleaned by checking the original data for duplicate records, system-missing values, out-of-range values and logical inconsistencies.

#### *Statistical analysis*

##### *Descriptive statistics*

Means, standard deviations and percentages were calculated for the participants' characteristics. We calculated medians, 25th and 75th percentile values for the EBRB of the first measurement.

##### *Test-retest reliability and construct validity*

To determine test-retest reliability and construct validity, the agreement between the two measurements was assessed at the individual item level. For all continuous items ( $n = 57$ ) the two-way, random-effects, single-measure intraclass correlation coefficient (ICC) was calculated. ICC were classified as 'excellent' ( $\geq 0.81$ ), 'good' (0.61–0.80), 'moderate' (0.41–0.60) or 'poor' ( $\leq 0.40$ )<sup>(24–26)</sup>. For categorical items ( $n = 5$ ) or items with a dichotomous scale ( $n = 33$ ), Cohen's kappa coefficient ( $\kappa$ ) was calculated. The classification of  $\kappa$  values was the same as for the classification of ICC. Because the calculation of ICC and  $\kappa$  values depends on the variability in answering categories, we also calculated the percentage agreement classified as 'excellent' (90–100%), 'good' (75–89%), 'moderate' (60–74%) or 'poor' (<60%). When an ICC or  $\kappa$  value was lower than or equal to 0.40/0.60/0.80, but the percentage agreement was equal to or higher than 60%/75%/90% respectively, we determined the classification according to the percentage agreement<sup>(27)</sup>. All statistical analyses were performed using SPSS version 18.0.

#### **Results**

##### *Participants' characteristics*

A total of 111 adolescents from six schools participated in the test-retest reliability study and twenty adolescents from the same six schools participated in the construct validity study. The characteristics of both study populations are presented in Table 1. For none of the characteristics was a significant difference between the study populations found. The dropout rate was 10% (13/124) in the test-retest reliability study and 0% in the construct validity study.

##### *Energy balance-related behaviours*

Table 2 shows descriptive statistics of the EBRB of the test-retest reliability study and the construct validity study, based on calculations from the data of the first measurement.

**Table 1** Characteristics of adolescents participating in the test–retest reliability and construct validity studies; DOiT (Dutch Obesity Intervention in Teenagers) questionnaire assessing energy balance-related behaviours. Data are presented as *n* and % except where indicated

	Test–retest reliability		Construct validity	
	<i>n</i> or Mean	% or sd	<i>n</i> or Mean	% or sd
No. of participants in study, <i>n</i>	111	–	20	–
Age (years), mean and sd	13.3	0.5	13.2	0.8
Male gender	61	55	50	50
Ethnicity*				
Western (both parents are Western)	85	77	16	80
Non-Western	26	23	4	20
Family status				
Traditional (mother and father)	83	75	16	80
Non-traditional	28	25	4	20

\*This classification is based on that of CBS Statistics Netherlands. Someone with a Western background is someone originating from a country in Europe (excl. Turkey), North America, Oceania, Indonesia or Japan. Someone with a non-Western background is someone originating from a country in Africa, South America, Asia (excl. Indonesia and Japan) or Turkey.

**Table 2** Energy-balanced related behaviours of the adolescents participating in the test–retest reliability (*n* 111) and construct validity studies (*n* 20), based on calculations from the data of the first measurement; DOiT (Dutch Obesity Intervention in Teenagers) questionnaire. Data are presented as median and 25th percentile, 75th percentile

	Test–retest reliability		Construct validity	
	Median	25th, 75th percentile	Median	25th, 75th percentile
Soft drink consumption (ml/d)	886	418, 1436	657	312, 1708
Fruit juice consumption (ml/d)	247	55, 670	327	166, 686
Consumption of sweets* (portions/d)	1	0, 2	1	1, 2
Snack consumption (portions/d)	1	0, 2	0	0, 1
Breakfast frequency (d/week)	7	4, 7	7	4, 7
Television viewing time (min/d)	137	94, 197	133	44, 210
Computer time (min/d)	120	56, 180	84	60, 197
Screen viewing time (min/d)	274	165, 358	244	120, 324
Active transport (min/school day)	26	16, 36	16	6, 28
Sport (in clubs) (min/d)	17	0, 43	26	11, 34
Sport (total)† (min/d)	43	21, 77	47	28, 67

\*Including candies, cookies, chocolates and ice cream.

†Including organised (in a club or at school) and unorganised sport activities.

### General findings

Results of the test–retest reliability and construct validity study for all questionnaire items of the DOiT questionnaire are presented in Table 3. For each item, the ICC or  $\kappa$  value and the percentage agreement are presented for both the test–retest reliability and construct validity. Table 4 shows a summary of the results per section of the questionnaire.

### Test–retest reliability

For the total population, thirty-two (34%) items had excellent test–retest reliability, forty-two (44%) items had good and twenty-one (22%) items had moderate test–retest reliability. No item had poor test–retest reliability. Most items with moderate scores were on consumption of sugar-containing beverages (five items on soft drinks and six items on fruit juices) and high-energy snacks and sweets (five items on snacks and two items on sweets). Concerning sugar-containing beverages, the ICC/ $\kappa$  values of ten out of twelve items on cartons/small bottles and glasses indicated moderate test–retest reliability. One item

on physical activity ('hours of after-school-time physical activity at school') showed moderate test–retest reliability. Fourteen items of the questionnaire showed low variability, resulting in ICC  $\leq$  0.60, but a high percentage agreement ( $\geq$ 90%).

To gain insight into possible differences between boys and girls, we also performed gender-specific analyses for the test–retest reliability. The analyses revealed some differences between boys and girls (presented in Supplementary Materials, Tables 1 and 2). In general, items showed values indicating lower test–retest reliability in girls. There were no items with poor test–retest reliability in the total population or boys only. In girls, 7% (*n* 7) of the items scored low on test–retest reliability. Four out of these seven items concerned yesterday's behaviour.

### Construct validity

For thirty-three (35%) items construct validity was excellent, eighteen (19%) items had good and twenty-three (24%) items moderate validity. For twenty-one (22%) items we found values indicating poor construct validity. Eight items

**Table 3** Agreement between questionnaires for the test–retest reliability study (*n* 111) and between questionnaires and interviews for the construct validity study (*n* 20), as indicated by intraclass correlation coefficient (ICC) or Cohen's kappa coefficient ( $\kappa$ ) and percentage agreement (% Agree); DOIr (Dutch Obesity Intervention in Teenagers) questionnaire assessing energy balance-related behaviours. Results are presented per questionnaire item

Item	Reliability		Validity	
	ICC/ $\kappa$	% Agree	ICC/ $\kappa$	% Agree
<b>Consumption of soft drinks</b>				
How many days a week do you drink soft drinks? (B1)	0.59	51	0.34	60
How many soft drinks do you drink on a school day? Cartons/small bottles (200 ml) (B2a)	0.74	67	0.12	35
How many soft drinks do you drink on a school day? Glasses (200 ml) (B2b)	0.45	46	0.47	35
How many soft drinks do you drink on a school day? Cans (330 ml) (B2c)	0.61	61	0.24	70
How many soft drinks do you drink on a school day? Bottles (500 ml) (B2d)	0.28	81	0.17	75
How many soft drinks do you drink on a weekend day? Cartons/small bottles (200 ml) (B3a)	0.52	67	0.16	55
How many soft drinks do you drink on a weekend day? Glasses (200 ml) (B3b)	0.63	33	0.34	45
How many soft drinks do you drink on a weekend day? Cans (330 ml) (B3c)	0.63	67	0.39	70
How many soft drinks do you drink on a weekend day? Bottles (500 ml) (B3d)	0.62	76	-0.04	55
How many soft drinks did you drink yesterday? Cartons or small bottles (200 ml) (B4a)	0.55	57	0.32	45
How many soft drinks did you drink yesterday? Glasses (200 ml) (B4b)	0.51	37	0.53	47
How many soft drinks did you drink yesterday? Cans (330 ml) (B4c)	0.57	75	0.53	80
How many soft drinks did you drink yesterday? Bottles (500 ml) (B4d)	0.58	77	0.46	60
<b>Consumption of fruit juices</b>				
How many days a week do you drink fruit juices? (B5)	0.75	42	0.41	25
How many fruit juices do you drink on a school day? Cartons/small bottles (200 ml) (B6a)	0.54	66	0.40	45
How many fruit juices do you drink on a school day? Glasses (200 ml) (B6b)	0.52	50	0.19	42
How many fruit juices do you drink on a school day? Cans (330 ml) (B6c)	0.36	90	0.00	84
How many fruit juices do you drink on a school day? Bottles (500 ml) (B6d)	0.45	79	-0.06	74
How many fruit juices do you drink on a weekend day? Cartons/small bottles (B7a)	0.36	73	-0.12	63
How many fruit juices do you drink on a weekend day? Glasses (200 ml) (B7b)	0.57	41	0.12	25
How many fruit juices do you drink on a weekend day? Cans (330 ml) (B7c)	0.36	88	0.00	83
How many fruit juices do you drink on a weekend day? Bottles (500 ml) (B7d)	0.40	79	0.00	79
How many fruit juices did you drink yesterday? Cartons or small bottles (200 ml) (B8a)	0.50	66	0.45	60
How many fruit juices did you drink yesterday? Glasses (200 ml) (B8b)	0.55	56	0.30	35
How many fruit juices did you drink yesterday? Cans (330 ml) (B8c)	0.57	94	*	100
How many fruit juices did you drink yesterday? Bottles (500 ml) (B8d)	0.41	86	-0.08	84
<b>Consumption of sweets</b>				
How many days a week do you eat sweets? (C1)	0.66	48	0.60	30
How many sweets do you eat on a school day? (C2)	0.71	49	0.21	10
How many sweets do you eat on a weekend day? (C3)	0.73	37	0.07	15
How many sweets did you eat yesterday? (C4)	0.59	47	0.37	15
What kind of sweets do you eat most? (C5)	0.60	68	0.14	33
<b>Consumption of snacks</b>				
How many days a week do you eat snacks? (C6)	0.50	31	-0.11	15
How many small snacks (chips/peanuts) do you eat on a school day? (C7a)	0.62	45	0.13	25
How many big snacks (hamburger/French fries) do you eat on a school day? (C7b)	0.58	59	-0.08	68
How many small snacks do you eat on a weekend day? (C8a)	0.53	36	0.44	35
How many big snacks do you eat on a weekend day? (C8b)	0.64	56	0.08	42
How many small snacks did you eat yesterday? (C9a)	0.49	56	0.25	50
How many big snacks did you eat yesterday? (C9b)	0.46	68	-0.07	78

Table 3 Continued

Item	Reliability		Validity	
	ICC/ $\kappa$	% Agree	ICC/ $\kappa$	% Agree
<b>Breakfast/meals</b>				
On how many days during the school week (Monday to Friday) do you eat breakfast? (D1)	0.83	84	0.91	75
On how many days in the weekend (Saturday and Sunday) do you eat breakfast? (D2)	0.64	86	0.63	85
What do you usually drink for breakfast? Nothing (D3a)	0.37	94	*	95
What do you usually drink for breakfast? Tea/coffee (D3b)	0.72	88	0.67	85
What do you usually drink for breakfast? Milk/yoghurt drink (D3c)	0.71	86	0.69	85
What do you usually drink for breakfast? Fruit juice (D3d)	0.67	89	0.16	60
What do you usually drink for breakfast? Liquid breakfast (D3e)	0.65	96	*	100
What do you usually drink for breakfast? Soft drink/energy drink (D3f)	0.54	94	0.57	85
What do you usually eat for breakfast? Nothing (D3g)	0.82	98	0.62	90
What do you usually eat for breakfast? Bread/crackers(D3h)	0.66	86	0.12	60
What do you usually eat for breakfast? Cereals (D3i)	0.68	88	0.48	85
What do you usually eat for breakfast? Croissant/coffee roll (D3j)	0.51	88	0.00	65
What do you usually eat for breakfast? Biscuit/sweet/snack (D3k)	*	99	*	100
Did you eat breakfast yesterday? (D4)	0.77	93	1.00	100
Did you eat lunch yesterday? (D5)	0.37	86	0.44	90
Did you eat dinner yesterday? (D6)	-0.02	95	*	100
<b>Screen behaviour</b>				
How long do you watch television on a school day? (min/d) (including watching DVD) (E1a)	0.66	44	0.52	35
How long do you watch television on a weekend day? (including watching DVD) (E1b)	0.65	35	0.52	15
How long did you watch television yesterday? (E2)	0.60	42	0.53	25
How long do you spend behind the computer for playing games and leisure activities on a school day? (E3a)	0.62	47	0.65	45
How long do you spend behind the computer for playing games and leisure activities on a weekend day? (E3b)	0.64	29	0.25	35
How long did you spend behind the computer for playing games and leisure activities yesterday? (E4)	0.43	30	0.73	35
<b>Active transport</b>				
How many days a week do you walk to school? (F1)	0.21	97	*	100
If you walk to school, how long does it take you to walk to school? (F2)	0.65	97	*	100
How many days a week do you bike to school? (F3)	0.78	94	1.00	100
If you bike to school, how long does it take you to bike to school? (F4)	0.91	64	0.68	45
Did you go to school in other ways than walking or biking in the last 5 school days? No (F5a)	0.32	88	0.46	90
Did you go to school in other ways than walking or biking in the last 5 school days? Yes, ... times by public transport (F5b)	0.86	96	0.00	95
Did you go to school in other ways than walking or biking in the last 5 school days? Yes, ... times by car/scooter/moped (F5c)	0.74	88	1.00	100
How did you go to school yesterday? By bicycle (F6a)	0.17	88	1.00	100
How did you go to school yesterday? Walking (F6b)	-0.01	96	1.00	100
How did you go to school yesterday? By public transport (F6c)	0.32	96	*	100
How did you go to school yesterday? By car/scooter/moped (F6d)	0.19	94	1.00	100
When you go somewhere else than school, how do you go there? By bicycle (F7a)	0.21	87	*	95
When you go somewhere else than school, how do you go there? Walking (F7b)	0.47	84	0.03	65
When you go somewhere else than school, how do you go there? By public transport (F7c)	0.22	95	*	100
When you go somewhere else than school, how do you go there? By car/scooter/moped (F7d)	0.58	83	0.24	70
<b>Physical activity</b>				
Do you participate in a sport at a sports club? (G1)	0.98	99	0.86	95
How many hours a week do you do this sport? (G3)	0.94	73	0.78	53
Do you participate in a second sport at a sports club? (G4)	0.79	95	0.69	90
How many hours a week do you do this second sport? (G6)	0.76	87	0.96	90
Do you participate in a third sport at a sports club? (G7)	*	99	*	100

**Table 3** Continued

Item	Reliability		Validity	
	ICC/ $\kappa$	% Agree	ICC/ $\kappa$	% Agree
How many hours a week do you do this third sport? (G9)	0.00	99	*	100
Do you participate in sports outside a sports club? (G10)	0.64	84	0.33	69
How many hours a week do you do these sports? (G11)	0.64	47	0.45	28
How long did you play sports yesterday? (at a sports club and outside a sports club) (G12)	0.63	37	-0.17	26
How important do you think it is to participate in sports? (from very important to not important at all) (G13)	0.75	85	0.07	45
What do you usually do during physical activity lessons at school? (H1)	*	96	*	85
Which of the following statements fits the most to the things you do during recess? (H2)	0.66	81	*	75
How many hours do you do after-school-time physical activities at school? (H3)	0.47	66	-0.06	50
<b>Having a job</b>				
Do you have an additional job at the moment? (I1)	0.91	96	0.83	95
I have a job – bringing around newspapers (I2a)	0.96	99	1.00	100
I have a job – stocking shelves at the supermarket (I2b)	1.00	100	*	100
I have a job – babysitting (I2c)	0.93	99	*	95
I have a job – cleaning (I2d)	0.66	99	*	100
I have a job – something else (I2e)	1.00	100	1.00	100
How many hours a week do you work? (I3)	0.97	89	0.97	95

\*Zero variance.

**Table 4** Overview of results of the test-retest reliability study and construct validity studies for all participants; DOIT (Dutch Obesity Intervention in Teenagers) questionnaire assessing energy balance-related behaviours

Items (section)	No. of items	Reliability (n 111)						Validity (n 20)					
		Range of ICC/ $\kappa$	Range of % Agree	Excellent	Good	Moderate	Poor	Range of ICC/ $\kappa$	Range of % Agree	Excellent	Good	Moderate	Poor
Consumption of soft drinks (B)	13	0.28–0.74	33–81	–	8	5	–	0.12–0.53	35–80	–	2	6	5
Consumption of fruit juices (B)	13	0.36–0.75	41–94	2	5	6	–	0.00–0.45	25–100	1	4	4	4
Consumption of sweets (C)	5	0.59–0.73	37–68	–	3	2	–	0.07–0.60	10–33	–	–	1	4
Consumption of snacks (C)	7	0.46–0.64	31–68	–	2	5	–	0.08–0.44	15–78	–	1	2	4
Breakfast/meals (D)	16	0.37–0.83	84–99	8	8	–	–	0.00–1.00	60–100	8	5	3	–
Screen behaviour (E)	6	0.43–0.66	29–47	–	4	2	–	0.25–0.73	15–45	–	2	3	1
Active transport (F)	15	0.17–0.91	64–96	9	6	–	–	0.00–1.00	45–100	12	1	2	–
Physical activity (G+H)	13	0.00–0.98	37–99	6	6	1	–	0.07–0.96	28–100	5	3	2	3
Having a job (I)	7	0.66–1.00	89–100	7	–	–	–	0.00–1.00	95–100	7	–	–	–
Overall	95	0.00–1.00	29–100	32 (34%)	42 (44%)	21 (22%)	–	0.00–1.00	10–100	33 (35%)	18 (19%)	23 (24%)	21 (22%)

ICC, intraclass correlation coefficient;  $\kappa$ , Cohen's kappa coefficient; % Agree, percentage agreement.

on high-energy snacks/sweets and nine items on sugar-containing beverages showed poor validity values. Most of these items concerned the amount of drinks or portions of food the adolescent had on a weekday or weekend day. Sixteen items showed low variability, resulting in  $ICC \leq 0.40$ , but a high percentage agreement ( $\geq 90\%$ ).

The study sample of the construct validity study was too small for gender-specific analyses.

## Discussion

The current study examined the test–retest reliability and construct validity of the DOiT questionnaire among 12- to 14-year-old Dutch adolescents attending pre-vocational education. The DOiT questionnaire, measuring EBRB, showed good test–retest reliability and moderate to good construct validity.

More than three-quarters of all items (78%) showed good to excellent test–retest reliability. The test–retest reliability appeared to be moderate for all other items. Most items with moderate scores were on consumption of soft drinks, fruit juices, sweets and snacks. Notable was the moderate test–retest reliability of the items concerning the consumption of sugar-containing beverages, especially the question on the amount of cartons/small bottles and glasses.

Fifty-four per cent of the items showed good to excellent construct validity, 24% of the items showed moderate construct validity and 22% of the items poor construct validity. Most items with poor validity concerned consumption of sugar-containing beverages (mainly the amount, i.e. cartons/small bottles and glasses) and high-energy snacks/sweets, which are the same questionnaire sections in which items often scored moderate on test–retest reliability.

There was a certain overlap in items scoring poor on construct validity and items scoring moderate on test–retest reliability. Especially recall of drinks and snacks, which are consumed throughout the day, seems to be difficult for adolescents. Behaviours such as screen behaviour in leisure time or consumption of breakfast may be more structured or stable or attached to specific parts of the day, and therefore easier to recall. Adolescents might also have had difficulties differentiating between the different packing sizes of drinks or between sweets and snacks, despite of examples given in the questionnaire and by the research assistant. Another explanation is that these concepts might be not defined explicitly enough in the questionnaire.

Therefore, we suggest that items regarding the consumption of sugar-containing beverages and high-energy snacks/sweets should be simplified and if used in future studies in the current form, the poor to moderate construct validity should be noted in interpreting research results based on these items. It should be considered that earlier research showed the effectiveness of the DOiT intervention in reducing the consumption of sugar-containing beverages<sup>(6)</sup>.

## Comparison with other studies

No comparable questionnaires and corresponding test–retest reliability and construct validity studies among adolescents aged 12–14 years were found for the broad range of EBRB assessed in the DOiT questionnaire. However, there are reliability and validity studies focusing on specific EBRB. It should be considered that validity is often assessed by different procedures than the method used in the current study. Because of the differences in methods and questionnaires, the results are difficult to compare.

Neuhouser *et al.*<sup>(28)</sup> examined the test–retest reliability and validity of the Beverage and Snack Questionnaire (BSQ) among young American adolescents (mean age 12.7 years). The test–retest reliability ranged from  $r = 0.62$  to  $r = 0.89$ . Validity coefficients comparing the BSQ with a 4 d food record ranged from  $r = 0.48$  to  $r = 0.87$ . In the DOiT questionnaire the items on sugar-containing beverages and high-energy snacks/sweets had slightly lower reliability ( $ICC/\kappa = 0.28–0.75$  and agreement = 31–94%) and validity ( $ICC/\kappa = 0.00–0.60$  and agreement = 10–100%). Unlike the DOiT questionnaire, the BSQ distinguishes between beverages and snacks/sweets consumed at school and not at school.

In a study by Hardy *et al.*<sup>(29)</sup> the reliability of the Adolescent Sedentary Activity Questionnaire (ASAQ) was examined in 11- to 15-year-old Australian adolescents. ICC values indicated good to excellent reliability among all students of grade 8 (mean age 13.3 years) for screen behaviour ( $ICC = 0.78$  for girls and  $ICC = 0.90$  for boys), while our study showed moderate to good reliability values for screen behaviour ( $ICC = 0.43–0.66$ ). A study by Chinapaw *et al.*<sup>(30)</sup> showed moderate reliability values on sedentary behaviour ( $ICC = 0.57$ ) as well as on moderate- and vigorous-intensity physical activity ( $ICC = 0.50–0.59$ ) assessed by the Activity Questionnaire for Adults and Adolescents (AQuAA) in Dutch adolescents aged 12–16 years. The construct validity of the AQuAA compared with assessment by accelerometry was poor (Spearman correlation coefficients =  $-0.21$  to  $0.23$ ). Also the WHO Health Behaviour in School-aged Children (HBSC) questionnaire and the International Physical Activity Questionnaire (IPAQ, short version) did not seem to be a valid instrument compared with a 7 d activity monitoring instrument for measuring physical activity in Norwegian adolescents aged 13–18 years (Spearman correlation coefficients =  $0.01–0.29$ ). The reliability was better in the WHO HBSC questionnaire ( $ICC = 0.71–0.73$ ) compared with the IPAQ ( $ICC = 0.10–0.62$ )<sup>(31)</sup>. Our study showed good to excellent test–retest reliability ( $ICC/\kappa = 0.00–0.98$  and agreement = 37–99%) and mixed values for the construct validity study on physical activity items ( $ICC/\kappa = 0.07–0.96$  and agreement = 28–100%).

## Strengths and limitations

Assessing a broad variety of EBRB is a major strength of the DOiT questionnaire. Besides, data were collected and

managed by standardised protocols. Nevertheless, there are also several limitations in our study. The sample size in the construct validity study was relatively small ( $n = 20$ ). Another limitation was the lack of a 'gold standard' in the construct validity study. Although comparison of the questionnaire with information from a cognitive interview was the most feasible and informative option in the present study, both measurement tools are self-reports and cognitive interviewing might lead to bias. We tried to minimise social desirability bias by accentuating the importance of accurate and honest answers. We restricted possible bias due to interpretation of the interview responses by processing of the interview data by a person other than the interviewer and according to a strict data entry protocol. An advantage of using interviews was to obtain additional information about the interpretation of the questions by the adolescents, based on which the DOiT questionnaire can be adapted for future use.

### Conclusions

The results of our study demonstrated good test–retest reliability and moderate to good construct validity of a majority of items from the DOiT questionnaire assessing EBRB in Dutch adolescents. All items with poor construct validity should be revised and tested again to improve the DOiT questionnaire for future use. Differentiating between consumption of sugar-containing beverages and high-energy snacks and sweets at school and out of school is another point to consider in a later version of the DOiT questionnaire.

### Acknowledgements

*Sources of funding:* This research received no specific grant from any funding agency in the public, commercial or non-for-profit sectors. *Conflicts of interest:* All authors declare to have no conflict of interests. *Authors' contributions:* A.S.S., M.J.M.C., F.v.N., J.B. and W.v.M. developed the measurement instrument. A.S.S. developed the study protocol. F.v.N. coordinated and supervised the data collection. F.v.N. and A.S.S. contributed to or supervised the data collection. E.H.C.J. conducted the data analyses under supervision of A.S.S. and M.J.M.C. E.H.C.J. and A.S.S. drafted the manuscript. All authors read and approved the final manuscript.

### Supplementary Materials

For Supplementary Materials for this article, please visit <http://dx.doi.org/10.1017/S1368980012005253>

### References

- Lobstein T, Baur L & Uauy R (2004) Obesity in children and young people: a crisis in public health. *Obes Rev* **5**, Suppl. 1, 4–104.
- World Health Organization (2003) *Diet, Nutrition and Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation. WHO Technical Support Series* no. 916. Geneva: WHO.
- World Health Organization (2004) *Global Strategy on Diet, Physical Activity and Health*. Geneva: WHO.
- Brown T & Summerbell C (2009) Systematic review of school-based interventions that focus on changing dietary intake and physical activity levels to prevent childhood obesity: an update to the obesity guidance produced by the National Institute for Health and Clinical Excellence. *Obes Rev* **10**, 110–141.
- Singh AS, Chinapaw MJ, Kremers SP *et al.* (2006) Design of the Dutch Obesity Intervention in Teenagers (NRG-DOiT): systematic development, implementation and evaluation of a school-based intervention aimed at the prevention of excessive weight gain in adolescents. *BMC Public Health* **6**, 304.
- Singh AS, Chinapaw MJ, Brug J *et al.* (2009) Dutch obesity intervention in teenagers: effectiveness of a school-based program on body composition and behavior. *Arch Pediatr Adolesc Med* **163**, 309–317.
- McDowell I (2006) *Measuring Health. A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press Inc.
- Steiner DL & Norman GR (2011) *Health Measurement Scales. A Practical Guide to Their Development and Use*. New York: Oxford University Press Inc.
- Montoye HJ, Kemper HCG, Saris WHM *et al.* (1996) *Measuring Physical Activity and Energy Expenditure*. Champaign, IL: Human Kinetics.
- Chinapaw MJ, Mokkink LB, van Poppel MN *et al.* (2010) Physical activity questionnaires for youth: a systematic review of measurement properties. *Sports Med* **40**, 539–563.
- Collins CE, Watson J & Burrows T (2010) Measuring dietary intake in children and adolescents in the context of overweight and obesity. *Int J Obes (Lond)* **34**, 1103–1115.
- Lubans DR, Hesketh K, Cliff DP *et al.* (2011) A systematic review of the validity and reliability of sedentary behaviour measures used with children and adolescents. *Obes Rev* **12**, 781–799.
- Van Nassau F, Singh AS, van Mechelen W *et al.* (2012) Exploring facilitating factors and barriers to the nationwide dissemination of a school-based weight gain prevention program 'DOiT': a study protocol. Submitted for publication.
- Booth ML, Okely AD, Chey TN *et al.* (2002) The reliability and validity of the Adolescent Physical Activity Recall Questionnaire. *Med Sci Sports Exerc* **34**, 1986–1995.
- Robinson TN (1999) Reducing children's television viewing to prevent obesity: a randomized controlled trial. *JAMA* **282**, 1561–1567.
- van Assema P, Brug J, Ronda G *et al.* (2001) 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 Diet* **14**, 377–390.
- van Assema P, Brug J, Ronda G *et al.* (2002) A short Dutch questionnaire to measure fruit and vegetable intake: relative validity among adults and adolescents. *Nutr Health* **16**, 85–106.
- Brug J, te Velde SJ, Chinapaw MJ *et al.* (2010) Evidence-based development of school-based and family-involved prevention of overweight across Europe: the ENERGY-project's design and conceptual framework. *BMC Public Health* **10**, 276.
- Singh AS, Vik FN, Chinapaw MJ *et al.* (2011) Test–retest reliability and construct validity of the ENERGY-child questionnaire on energy balance-related behaviours and their potential determinants: the ENERGY-project. *Int J Behav Nutr Phys Act* **8**, 136.

20. van Stralen MM, te Velde SJ, Singh AS *et al.* (2011) European Energy balance Research to prevent excessive weight Gain among Youth (ENERGY) project: design and methodology of the ENERGY cross-sectional survey. *BMC Public Health* **11**, 65.
21. RIVM (2005) Lokale en Nationale Monitor Gezondheid. <https://www.monitoregezondheid.nl/jeugdindicatoren.aspx> (accessed September 2009).
22. Kirshner B & Guyatt G (1985) A methodological framework for assessing health indices. *J Chronic Dis* **38**, 27–36.
23. McMinn AM, van Sluijs EM, Harvey NC *et al.* (2009) Validation of a maternal questionnaire on correlates of physical activity in preschool children. *Int J Behav Nutr Phys Act* **6**, 81.
24. Landis JR & Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.
25. McGraw KO & Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* **1**, 30–46.
26. Nunnally JC & Bernstein IH (1994) *Psychometric Theory*. New York: McGraw-Hill.
27. Saelens BE, Frank LD, Auffrey C *et al.* (2006) Measuring physical environments of parks and playgrounds: EAPRS instrument development and inter-rater reliability. *J Phys Act Health* **3**, Suppl. 1, S190–S207.
28. Neuhouser ML, Lilley S, Lund A *et al.* (2009) Development and validation of a beverage and snack questionnaire for use in evaluation of school nutrition policies. *J Am Diet Assoc* **109**, 1587–1592.
29. Hardy LL, Booth ML & Okely AD (2007) The reliability of the Adolescent Sedentary Activity Questionnaire (ASAQ). *Prev Med* **45**, 71–74.
30. Chinapaw MJ, Sloomaker SM, Schuit AJ *et al.* (2009) Reliability and validity of the Activity Questionnaire for Adults and Adolescents (AQuAA). *BMC Med Res Methodol* **9**, 58.
31. Rangul V, Holmen TL, Kurtze N *et al.* (2008) Reliability and validity of two frequently used self-administered physical activity questionnaires in adolescents. *BMC Med Res Methodol* **8**, 47.