Food Research 9 (2) : 101 - 110 (April 2025)
Journal homepage: https://www.myfoodresearch.com

FOOD RESEARCH

RESEARCH PAPER

The single-item Healthy Diet Scale: validity and reliability

[1]van Oostrom, E.C., [1]Hendriksen, P.A., [1]van der Weij, B.R.C., [1]Boogaard, A.S., [1]Mulder, K.E.W., [1]Verheul, M.C.E., [1]Tan, S., [1]Kraneveld, A.D., [2,3]Vlieg-Boerstra, B., [1,4]Garssen, J. and [1,5,6,*]Verster, J.C.

[1]*Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584CG Utrecht, The Netherlands*
[2]*OLVG Hospital, Department of Paediatrics, Amsterdam, the Netherlands*
[3]*Rijnstate Allergy Center, Rijnstate, Arnhem, the Netherlands*
[4]*Danone Global Research and Innovation Centre, 3584CT Utrecht, The Netherlands*
[5]*Centre for Mental Health and Brain Sciences, Swinburne University, Melbourne, VIC 3122, Australia*
[6]*Cognitive Neurophysiology Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, D-01307 Dresden, Germany*

Article history:
Received: 25 October 2024
Received in revised form: 21 January 2025
Accepted: 30 January 2025
Available Online: 11 March 2025

Keywords:
Healthy diet scale,
Sleep,
Immune fitness,
Mood,
Quality of life

DOI:
https://doi.org/10.26656/fr.2017.9(2).194

Abstract

The single-item Healthy Diet Scale (HDS) was developed to estimate the percentage of daily diet that consumers regard as healthy. The aim of the presented studies was to demonstrate its validity and reliability. To evaluate its validity, in Study 1, N = 108 Dutch young adults completed a survey including the HDS (with picture examples) and the multiple-item Start the Conversation diet scale (STC). To examine the reliability of the HDS, a same-day test-retest assessment was conducted. In Study 2, an online survey among N = 154 Dutch students compared the HDS with and without picture examples. In Study 3, among N = 302 Turkish students, the HDS (without picture examples) was compared with the STC. In Study 4, among N = 37 Dutch and N = 83 English language students, the test-retest reliability of the HDS (without picture examples) was assessed for a longer time interval. Study 5, among N = 100 Dutch students, evaluated the correlation of the HDS (without picture examples) with various health outcomes. In Study 1, the analysis revealed a significant correlation between the HDS with picture examples and the STC. The intraclass correlation (ICC) between the test and retest assessment of 0.982 suggested excellent agreement between the two assessments. Study 2 revealed no significant difference in the HDS with and without example pictures. Study 3 revealed that the outcome of the HDS without picture examples correlated significantly with the STC. Study 4 revealed that the 11-day test-retest reliability of the HDS without picture examples was good to excellent. Study 5 revealed a significant association between the HDS without picture examples and immune fitness, sleep quality, insomnia, happiness and quality of life, but not with mood. Taken together, the HDS is a valid and reliable scale to estimate the percentage of daily diet that consumers perceive as healthy.

1. Introduction

A healthy diet refers to a diet in which beneficial nutrients, such as vitamins, fibers, and minerals, are provided in appropriate proportions, and is one of the pillars of attaining a healthy lifestyle (Cena and Calder, 2020). A healthy diet is essential to maintain an adequate immune fitness, i.e., the ability to initiate effective responses when pathogens such as bacteria and viruses encounter the body (Di Giosia *et al.*, 2022). In addition, a healthy diet is essential for a resilient immune system and as such supports the prevention of non-communicable diseases (Cena and Calder, 2020).

Although there are biomarkers that refer to components of dietary intake (e.g., plasma phospholipid fatty acid composition, the assessment of nutrients, or body composition), there are no biomarkers available that accurately represent the overall concept of a healthy or unhealthy diet. Therefore, one has to rely on scales and questionnaires that are completed by the individuals under investigation (Shim *et al.*, 2014). To evaluate

*Corresponding author.
Email: *j.c.verster@uu.nl*

eISSN: 2550-2166
Published by Rynnye Lyan Resources

dietary intake, food frequency questionnaires (FFQs) are commonly used (Shim *et al.*, 2014). These are elaborate questionnaires assessing the quantity and frequency of food intake. Unfortunately, it is very time-consuming to complete a FFQ and often this is considered as a burden to those who have to complete them. The increased patient burden might result in lower response and/or completion rates. In addition, FFQs yield less detailed information than alternative used methods. Alternatively, 24 hrs dietary recall methods or 3 to 7 days food diaries may be used to collect this data (Shim *et al.*, 2014). Using this methodology, the individual under investigation writes down all food products, including brands and serving sizes, they consumed during the past 24 hrs. Disadvantages of FFQs, 24 hrs recalls, and food diaries are recall bias (i.e., people will forget to list food items), that the outcomes are not readily available, and that they are difficult to use by health care professionals without dietary knowledge. Extensive calculations are required to determine the daily intake of individual nutrients. The decision to what extent the diet can be considered balanced and healthy then depends on the interpretation of the researcher that evaluates the nutrient scores. Taken together, FFQs and 24 hrs dietary recall may be viewed as time consuming and sub-optimal methodologies to infer to what extent a diet can be regarded as healthy, and there are no standardized guidelines available for this purpose.

A literature search revealed that there are several scales and questionnaires available that assess dietary habits, including the Food Habits Questionnaire (Glasgow *et al.*, 1996), the 23-item Rate Your Plate (Gans *et al.*, 2000), the Healthy Diet Index (Lindström *et al.*, 2021), and the 8-item Start The Conversation (STC) diet scale (Paxton *et al.*, 2011). Of these, the STC aims to assess healthful and unhealthful dietary behaviors and as such can be considered the best option to assess 'healthy diet'. The STC has been used successfully in previous research examining the relation-ship between attaining a healthy diet, physical activity and BMI (Ferrer *et al.*, 2016), the association of healthy diet with social support and mental health in US military veterans (Hoerster *et al.*, 2016), to investigate the usefulness of the STC in screening for overweight and obesity status in Hispanic women (McVey *et al.*, 2021), to evaluate healthy diet in relation to other health behaviors, mental health, and well-being during the COVID-19 pandemic (Shillington *et al.*, 2021), and to identify factors leading to unhealthy diet choices among international students (Alshehri *et al.*, 2023).

It is fundamental that a questionnaire reliably measures the concept under investigation. In the current study, the aim is to estimate the percentage of diet that can be considered as healthy, using a single-item measuring scale. It is unclear whether this can be inferred by interpreting a combination of nutrition values obtained from FFQs. Also, for multiple food-item questionnaires, it is unsure if the combination of items comprises all facets of the overall concept 'healthy diet'. Moreover, the outcome of these questionnaires is the sum score of the individual items, treating each individual item as an equally important determinant of a healthy diet, which is unlikely to be correct. In this context, the single-item Healthy Diet Scale (HDS) was developed (Figure 1).

Figure 1. The Healthy Diet Scale (HDS) with picture examples (Van Oostrom *et al.*, 2022).

The most important advantages of a single-item global assessment of a concept (in this case, a healthy diet) is that this single score approach evaluates the entire constellation of what constitutes the concept, instead of limiting the concept to the sum of the selected items that are part of a multiple item questionnaire (U.S. Food and Drug Administration [US FDA], 2009; Verster *et al.*, 2020; Verster *et al.*, 2021), and that it is quick and easy to use. The aim was to develop a single-item scale with a readily available and meaningful outcome that can be interpreted by any clinician or researcher without the need of any dietary knowledge. The HDS assesses what percentage of the past month diet that participants

https://doi.org/10.26656/fr.2017.9(2).194

consider healthy. The HDS was developed to be used in countries with a Western diet. Therefore, pictures and examples of healthy and unhealthy Western food items accompanied the rating scale to facilitate a more accurate response. The examples of healthy and unhealthy food were selected after thorough discussions between the authors (J.V. and B.V.) and encompass the most important food groups in a Western diet, i.e., fruits, vegetables, fat, protein sources, carbohydrates and drinks. The decision to include items was based on scientific literature, including the NOVA classification system of the Food and Agriculture Organization (FAO) of the United Nations (Monteiro *et al.*, 2019). The HDS aims to be a standard scale to assess perceived healthy diet status in clinical practice and research. The aim of the studies presented here was to evaluate the validity and reliability of the HDS. In the first study, the HDS with example pictures was evaluated. In 4 subsequent studies, the HDS without picture examples was evaluated. The use of the HDS without picture examples is not limited to Western countries. Since no food item examples are shown, it is hypothesized that it could be used across countries and cultures to assess the percentage of daily diet that consumers consider healthy.

2. Materials and methods

Five studies were conducted to develop the Healthy Diet Scale and evaluate its validity and reliability.

2.1 Study 1

In December 2021, N = 108 Dutch young adults completed two paper-pencil surveys. Participants were recruited among students of the department of pharmaceutical sciences of Utrecht University, The Netherlands. Participants could complete the surveys in the English or Dutch language. The study was reviewed and approved by the Science-Geo Ethics Review Board of Utrecht University (protocol code: S-21525, date of approval: November 21, 2021). All participants provided written informed consent and received 20-euro reimbursement. Demographic data (age and sex) were collected and diet was assessed with the HDS and STC. Approximately 30 minutes after completion of the survey, the HDS was again completed in order to evaluate its test-retest reliability. To prevent participants from actively memorizing the HDS outcome they provided in the first survey, they were distracted by other tasks (the collection of a saliva sample) between the two assessments. In addition, although subjects were aware that a second survey had to be completed, they were not aware that a retest of the HDS was included in the second survey.

The STC comprises 8 items on food intake (Paxton *et al.*, 2011). Both healthy and unhealthy food items are included. The frequency of dietary intake of each item is scored as 0, 1, or 2. A higher sum score of the 8 items represents an unhealthier diet. The HDS is a single-item rating scale to estimate the percentage of past month's diet that they perceive as healthy (Figure 1). Participants could rate the percentage of healthy diet on a scale ranging from 0% to 100%, in steps of 10%. Pictures and examples of healthy and unhealthy food items accompany the rating scale to facilitate a more accurate response. Copyright for the use of the selected pictures was obtained from Depositphotos Inc. (New York, USA).

Statistical analyses were conducted with SPSS (IBM IBM Corp., Statistics for Windows, Version 29.0, USA). Mean and standard deviation (SD) were computed for all variables, and normality of the data distribution was determined by visual inspection and the Kolmogorov–Smirnov test. The data was not normally distributed, and therefore nonparametric statistical tests were applied.

To evaluate the validity of the HDS, the Spearman's correlation was computed between the HDS and STC diet scale. The correlation was considered statistically significant if $p < 0.05$. To determine test-retest reliability of the HDS, the Spearman's correlation was computed between the test and retest assessment. The correlation was considered statistically significant if $p < 0.05$, and high test-retest reliability if $r > 0.8$ (Cohen, 1988). In addition, an average-measures, absolute-agreement, 2-way mixed-effects model was used to calculate the intraclass correlation (ICC) between the test and retest assessment of the HDS. The reliability of the ICC is determined by its 95% confidence interval (CI). If the 95% CI values are less than 0.5, this is indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability (Koo and Li, 2016). The 95% limits of agreement method by Bland and Altman were applied to confirm reliability (Bland and Altman, 1999). The difference score (DIFF) of the test and retest outcomes and the corresponding standard deviation (SDDIFF) were computed. According to the 95% limits of agreement method, there is agreement between the assessments if 95% of the DIFF score lies between (DIFF - 1.96 × SDDIFF) and (DIFF + 1.96 × SDDIFF). A Bland-Altman plot was composed, plotting DIFF scores for each participant against the mean score. If the datapoints are close to the line of equality (zero) and 95% of the data points lie between the lower and upper limit of the 95% limits of agreement interval, this implies high agreement between the two assessments (i.e., a high test-retest reliability for the HDS).

https://doi.org/10.26656/fr.2017.9(2).194

2.2 Study 2

The aims of Study 2 were (1) to compare the HDS with example pictures with the HDS without example pictures, and (2) to examine a possible impact of education level on understanding and interpretation of the HDS scale. To this extent, a paper-pencil survey was conducted among N = 154 Dutch young adults, aged 18 to 30 years old. The study was reviewed and approved by the Science-Geo Ethics Review Board of Utrecht University (protocol code: S-22743, date of approval: May 20, 2022). All participants provided written informed consent. The survey collected data on age and sex, and highest completed education level. First, participants completed the HDS without picture examples (Figure 2).

What did you eat last month?

Estimate the percentage of healthy food of your daily diet:	……..%

Answer possibilities: 0 % (unhealthy), 10 %, 20 %, 30 %, 40 %, 50 % (mixed), 60 %, 70 %, 80 %, 90 %, 100 % (healthy)

Figure 2. The Healthy Diet Scale without picture examples.

On the next survey page, participants were instructed to answer the HDS again but now with example pictures provided (Figure 1). Participants were instructed not to alter their previous answer (the HSD without examples). A subsequent question asked whether the picture examples were helpful to answer the HDS. Answer possibilities were (1) strongly disagree, (2) somewhat disagree, (3) neutral, (4) somewhat agree, and (5) strongly agree. It was further asked how much time it took to complete the HDS with picture examples. Finally, it was asked how much effort it took to complete the HDS wit picture examples. Answer possibilities were (1) no effort (I could immediately answer the question), (2) some effort, (3) considerable effort, and (4) great effort (It was difficult to answer the question; it took considerable time to think about the correct answer).

Statistical analyses were conducted with SPSS (IBM Corp., Statistics for Windows, Version 29.0, USA). Mean and standard deviation (SD) were computed for all variables, and normality of the data distribution was determined by visual inspection and the Kolmogorov–Smirnov test. The data was not normally distributed, and therefore nonparametric statistical tests were applied.

The outcome of the HDS with and without pictures was compared using the Related-Samples Wilcoxon Signed Rank test. The difference was considered significant if $p < 0.05$. In addition, a Spearman's correlation was computed. The correlation was considered statistically significant if $p < 0.05$. An average-measures, absolute-agreement, 2-way mixed-effects model was used to calculate the intraclass correlation (ICC) between the HDS with and without picture examples. Using the categorization of Statistics Netherlands (Centraal Bureau Voor de Statistiek, 2019), participants were categorized as lower or higher educated. Outcomes of lower and higher educated participants were compared using the Independent-Samples Mann-Whitney U test. Differences between the groups were considered significant if $p < 0.05$.

2.3 Study 3

The aim of Study 3 was to validate the HDS (without picture examples) by comparing its outcome with the STC. To this extent, an online survey was conducted among N = 302 Turkish University students (Hendriksen *et al.*, 2024). The study was reviewed and approved by the Science-Geo Ethics Review Board of Utrecht University (protocol code: S-23525c, date of approval: May 10, 2023). All participants provided electronic informed consent. Demographic data (age and sex) were collected and the HDS (without picture examples) and STC were completed.

Statistical analyses were conducted with SPSS (IBM Corp., Statistics for Windows, Version 29.0, USA). To evaluate the validity of the HDS (without picture examples), the Spearman's correlation was computed between the HDS and STC diet scale. The correlation was considered statistically significant if $p < 0.05$.

2.4 Study 4

The aim of Study 4 was to examine the test-retest reliability of the HDS (without picture examples), with a longer interval between the test and retest assessment. To this extent, an online survey was conducted among N = 120 students of Utrecht University. The sample comprised Dutch students and international students. The Dutch students completed the survey in the Dutch language, the international students completed the survey in the English language. The study was reviewed and approved by the Science-Geo Ethics Review Board of Utrecht University (protocol code: S-23044, date of approval: June 21, 2023). All participants provided written informed consent. Demographic data (date of birth and sex) were collected to match the test and retest assessments, and the HDS (without picture examples) was completed.

Statistical analyses were conducted with SPSS (IBM Corp., Statistics for Windows, Version 29.0, USA). To determine test-retest reliability of the HDS, the same analyses were conducted as described for Study 1. The analyses were conducted separately for the Dutch and English language sample.

https://doi.org/10.26656/fr.2017.9(2).194

2.5 Study 5

The aim of Study 5 was to evaluate to what extent attaining a healthy diet, assessed with the HDS (without picture examples), is associated with health outcomes, including immune fitness, sleep, mood, and quality of life. To this extent, an online survey was conducted among N = 100 students of Utrecht University. The study was reviewed and approved by the Science-Geo Ethics Review Board of Utrecht University (protocol code: S-23040, date of approval: June 27, 2023). All participants provided electronic informed consent. Demographic data (age and sex) were collected and the HDS (without picture examples) was completed. Immune fitness, i.e. the body's capacity to respond to health challenges (such as infections) by activating an appropriate immune response, was assessed with a single -item scale ranging from 0 (poor) to 10 (excellent) (Verster *et al.*, 2022; Verster *et al.*, 2023). Sleep quality was measured using a single-item scale, ranging from 0 (very poor) to 10 (excellent) (Donners *et al.*, 2015). Sleep was evaluated with the Regensburg Insomnia Scale (RIS) (Crönlein *et al.*, 2013). The 10 items of the RIS are scored on a 5-point Likert scale, and summed to a total score, ranging from 0 to 40. Higher RIS scores reflect greater insomnia complaints. Quality of life was assessed on a scale ranging from very poor (score 0) to excellent (score 10) (Verster *et al.*, 2023). Mood items (stress, anxiety, depression, fatigue, loneliness, hostility, and happiness) were assessed on single-item scales ranging from absent (score 0) to extreme (score 10) (Verster *et al.*, 2021; Verster *et al.*, 2023). All assessments were made for the past 6 months.

Statistical analyses were conducted with SPSS (IBM Corp., Statistics for Windows, Version 29.0, USA). Spearman's correlations were computed between the outcome of the HDS (without picture examples) and the other study outcomes. Correlations were considered statistically significant if $p < 0.05$ and $p < 0.007$ for mood items (applying a Bonferroni's correction for multiple comparisons).

3. Results

3.1 Study 1

A total of 108 Dutch young adults completed the study. They had a mean (SD) age of 21.5 (2.6) years old. The majority of the sample (71.3%) were women. Mean (SD) scores on the HDS and STC were 68.1% (14.3) and 6.4 (2.1), respectively. The distribution of the HDS scores is shown in Figure 3. No significant sex differences were found for scoring on the HDS and STC.

Correlations between the HDS and STC items are summarized in Table 1. A significant correlation was found between the HDS and STC sum score ($r = -0.577$, $p < 0.001$), see Figure 3. In addition, most STC items correlated significantly with the HDS score. The most robust correlations were found for the number of servings of vegetables per day and the frequency of eating fast-food per week (see Table 1).

The mean (SD) test 68.1±14.3%) and retest score (68.1±14.1%) of the HDS did not statistically differ from each other, and a high correlation was found between the two assessments ($r = 0.984$, $p < 0.001$). The intraclass

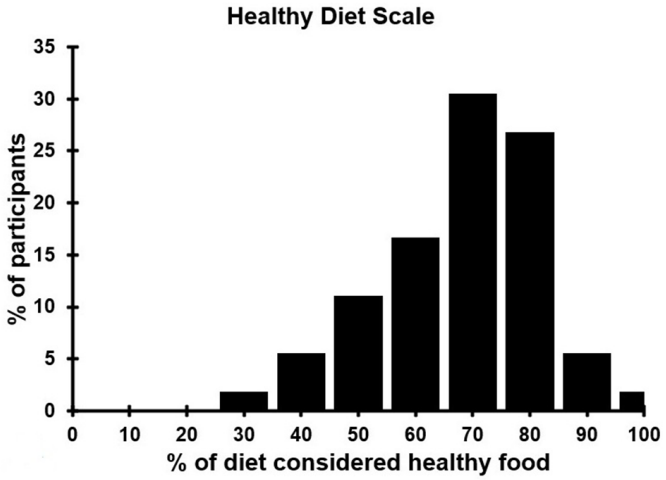

Figure 3. Distribution of Healthy Diet Scale score (with picture examples).

Table 1. Correlations of the HDS with STC items.

STC Item	r	*p*-value
How many times a week did you eat fast-food meals or snacks?	-0.345	<0.001 *
How many servings of fruit did you eat each day?	-0.314	<0.001 *
How many servings of vegetables did you eat each day?	-0.377	<0.001 *
How many regular sodas or glasses of sweet tea did you drink each day?	-0.219	0.023
How many times a week did you eat beans (like pinto or black beans), chicken, or fish?	-0.032	0.740
How many times a week did you eat regular snack chips or crackers (not low-fat)?	-0.310	<0.001 *
How many times a week did you eat desserts and other sweets (not the low-fat kinds)?	-0.250	0.009
How much margarine, butter, or meat fat do you use to season vegetables or put on potatoes, bread, or corn?	-0.202	0.036
STC sum score	-0.577	<0.001 *

Spearman's correlations between the HDS and STC items were considered significant if $p < 0.00625$ (after Bonferroni's correction).

https://doi.org/10.26656/fr.2017.9(2).194

correlation (ICC) between the test and retest assessments of the HDS was 0.991 (95% CI lower limit: 0.986, 95% CI upper limit: 0.994), indicating excellent reliability (Koo and Li, 2016).

In addition, the Bland-Altman method (Bland and Altman, 1999) was applied to further investigate the test-retest reliability of the HDS. The analysis yielded a mean (SD) difference score of 0.0 (2.7) with 95% limits of agreement ranging from -5.36 to 5.36. Only 5 out of 108 subjects (4.6%) revealed a different score on the test and retest assessment that was outside the 95% limits of agreement interval, suggesting that there was high agreement among both assessments. The Bland-Altman plot is shown in Figure 4.

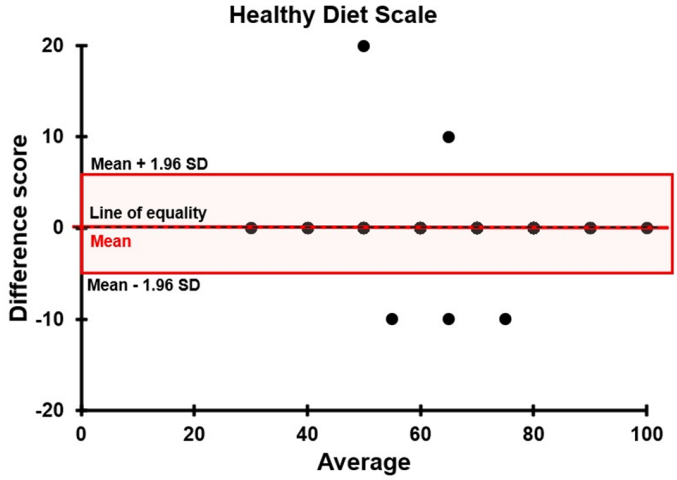

Figure 4. Bland-Altman plot for the Healthy Diet Scale (with picture examples). The line of equality (dotted black line), mean difference (red color line) and 95% limits of agreement (red box) are shown. Values on the y-axis represent the difference score between the test and retest assessments. Values on the x-axis represent the average score of the test and retest assessments.

3.2 Study 2

Study 2 aimed to compare the HDS with and without example pictures. N = 154 Dutch young adults participated in the study. Their mean (SD) age was 21.7 (3.2) years old, and 65.5% of the sample was female. The mean (SD) of the HDS without pictures was 59.1% (14.8%) and the mean (SD) of the HDS with picture examples was 60.2% (14.6%). The difference was not statistically significant ($p = 0.417$), and the correlation between the two assessments was significant ($r = 0.660$, $p < 0.001$). The intraclass correlation (ICC) was 0.777 (95% CI lower limit: 0.693, 95% CI upper limit: 0.837), indicating moderate to good correspondence between the two assessments (Koo and Li, 2016).

Of the sample, 62.7% had a higher education level. Compared to participants with a lower education level, participants with a higher education level reported a higher percentage of healthy diet for the HDS without picture examples (56.6% vs 60.6%, respectively, $p = 0.118$) and the HDS with picture examples (57.0% vs 61.9%, respectively, $p = 0.034$). No significant difference ($p = 0.676$) was found in HDS (with picture examples) completion time between lower (56.1 sec) and higher educated participants (38.4 sec). No significant differences between lower and higher educated participants were found with regard to the extent the example pictures helped completing the HDS.

Only a minority of the sample reported that the sample pictures were somewhat (31.2%) or very helpful (11.7%) to answer the HDS. A majority of the sample (70.8%) reported that it took no effort to complete the HDS with example pictures. In conclusion, the HDS with and without picture examples perform almost equally well, in both lower and higher educated individuals

3.3 Study 3

A total of 302 Turkish university students completed the study. They had a mean (SD) age of 22.1 (1.9) years old. The majority of the sample (70.9%) were women. Mean (SD) scores on the HDS and STC were 54.0% (25.8) and 8.4 (2.5), respectively. A significant correlation was found between the HDS and STC sum score ($r = -0.240$, $p < 0.001$).

3.4 Study 4

A total of 120 students (75% female) participated in the study. Their mean (SD) age was 20.5 (1.6) years old. N = 37 students completed the survey in English language (international students) and N = 83 students (Dutch students) completed the survey in Dutch language. English. For the English language sample, the mean (SD) test-retest interval was 11.0 (3.5) days (range: 7 to 14 days). The mean±SD test (67.8%±11.8%) and retest score (67.8%±14.0%) of the HDS did not statistically differ from each other. A high Spearman's correlation was found between the test and retest assessments ($r = 0.895$, $p < 0.001$). The intraclass correlation (ICC) between the test and retest assessments of the HDS was 0.958 (95% CI lower limit: 0.918, 95% CI upper limit: 0.978), indicating excellent reliability (Koo and Li, 2016). The Bland-Altman analysis yielded a mean (SD) difference score of 0.0 (5.3) with 95% limits of agreement ranging from -10.33 to 10.33. None of the 37 subjects (0.0%) had a different score on the test and retest assessment that was outside the 95% limits of agreement interval, suggesting that there was high agreement among both assessments (Bland and Altman, 1999).

For the Dutch language sample, the mean (SD) test-retest interval was 10.9 (3.5) days (range: 8 to 16 days). The mean±SD test (65.3%±13.3%) and retest score

https://doi.org/10.26656/fr.2017.9(2).194

(67.7%±14.3%) of the HDS statistically differed from each other (p = 0.026). However, the absolute difference has no clinical relevance on a scale ranging from 0% to 100%. A high Spearman's correlation was found between the test and retest assessments (r = 0.772, p < 0.001). The intraclass correlation (ICC) between the test and retest assessments of the HDS was 0.862 (95% CI lower limit: 0.783, 95% CI upper limit: 0.911), indicating good reliability (Koo and Li, 2016). The Bland-Altman analysis yielded a mean (SD) difference score of 2.4 (9.4) with 95% limits of agreement ranging from -16.02 to 20.82. Only 4 out of 83 subjects (4.8%) had a different score on the test and retest assessment that was outside the 95% limits of agreement interval, suggesting that there was high agreement among both assessments (Bland and Altman, 1999).

3.5 Study 5

A total of 100 Dutch students (83% female) participated in the study. Their mean (SD) age was 20.4 (1.8) years old. Their mean (SD) HDS score was 64.2% (14.1). The mean (SD) of outcome measures and their Spearman's correlation with the HDS (without picture examples is shown in Table 2. Attaining a healthy diet was significantly associated with a better immune fitness, improved sleep, and a greater quality of life and happiness. The HDS score did not significantly correlate with the other mood items.

4. Discussion

The five studies presented here showed that HDS is a quick, reliable and valid assessment tool to estimate the percentage of daily diet that is perceived as healthy. The studies demonstrated that the HDS performs equally well as the multiple-item STC diet scale, its reliability was confirmed via ICC, and associations were shown with immune fitness, sleep quality and insomnia, and quality of life. The HDS is a welcome addition to the research methodology in the field of nutrition. In contrast to elaborate FFQs, 24 hrs recall, or multiple-item questionnaires, no calculations or recoding are needed once the HDS is completed. It allows a real-time assessment, with a readily available outcome, that is easy to interpret.

The HDS and overall STC score correlated significantly with each other, suggesting that the HDS is a valid scale to assess healthy diet. Most pronounced were the correlations with the STC items on vegetables and fast-food meals and snacks consumption (Table 1), which are clear examples of healthy and unhealthy food, respectively. The fact that the strength of the correlations was modest, i.e. an r between -0.3 and -0.4, is most likely caused by the fact that the item single out food components that are part of but not fully comprise the concept of a healthy diet. Also, there are limited answering possibilities of the STC items. For example, for the number of fast-food meals and snacks a week one could choose between the answers (1) less than one time, (2) one to three times, and (3) four or more times. The 3rd answer possibility (i.e., four or more times) does not allow differentiation between participants (i.e., those who consume 25 snacks a week are treated equally as those who consume 4 snacks a week). This was also the case for other STC items.

The reliability of the HDS was evaluated by comparing the test and retest assessments. The Spearman's correlation, ICC, and the Bland-Altman limits of agreement method consistently suggest that the HDS has a good to excellent test-retest reliability. In Study 1 the test and retest assessment were made on the same day. This was necessary, because the test-retest assessment of the HDS was incorporated in a larger survey that aimed to evaluate test-retest assessments of

Table 2. Study outcomes and their correlation with the HDS (without picture examples).

Variable	Mean (SD)	r	*p*-value
Immune fitness	7.2 (1.4)	0.266	0.008*
Sleep quality	6.9 (1.4)	0.226	0.024*
RIS insomnia score	10.8 (4.8)	-0.257	0.010*
Quality of life	7.1 (1.3)	0.323	0.001*
Stress	6.8 (1.8)	-0.074	0.464
Anxiety	4.4 (2.9)	-0.052	0.607
Depression	3.1 (3.0)	-0.175	0.081
Fatigue	6.4 (2.4)	-0.164	0.103
Loneliness	3.8 (2.7)	-0.089	0.377
Hostility	1.8 (2.2)	-0.010	0.923
Happiness	6.8 (1.5)	0.283	0.004*

*Significant correlations with the HDS ($p < 0.05$, and $p < 0.007$ for mood items, applying a Bonferroni's correction for multiple comparisons).
RIS: Regensburg Insomnia Scale.

https://doi.org/10.26656/fr.2017.9(2).194

additional scales, including mood items. Due to day-to-day fluctuations in mood, these items required a same day test and retest assessments. Previous research has successfully applied a same day test and retest assessments (Mesquita *et al.*, 2013; Paiva *et al.*, 2014). It is very unlikely that participants were actively memorizing their answer to the first HDS assessment in order to reproduce it on the retest assessment. Participants were not aware that there was a retest session, and the time between the assessments was filled with another activity. The test-retest reliability was confirmed in Study 4, for Both Dutch and English language versions of the HDS without picture examples, with an average of 11 days between the test and retest assessment.

There is an ongoing discussion about which specific food items should be considered as healthy or unhealthy. As a result, there is no consensus among experts. For example, there is discussion about the classification of butter versus margarine, or meat. The example items for the HDS were chosen by the authors (J.V. and B.V.), based on scientific literature, including the NOVA classification system of the Food and Agriculture Organization (FAO) of the United Nations (Monteiro *et al.*, 2019). Although these encompass the most important food groups in a Western diet, i.e., fruits, vegetables, fat, protein sources, carbohydrates and drinks, others might disagree with our classification. The given examples are related to a Western diet and are not representative of other cultures. However, also between Western countries there is no consensus in guidelines about which specific food items should be considered healthy or unhealthy, and reasons for classification may vary (e.g., processed food, salt content). Given this, it is unlikely that the HDS with picture examples can be administered successfully in cultures and geographical regions with very different eating behaviors, such as Asia. Therefore, it was decided to further develop the HDS without picture examples (Studies 2 to 5). For future research, it is recommended to use the HDS without picture examples.

A limitation of the presented studies is that these were conducted in relatively healthy young adults. This may also explain the relatively high percentages of a healthy diet that were reported by the participants. According to the Dutch Food Consumption Survey 2019 -2021, the proportion of adults (18 – 69 years old) that meets the recommended daily amounts of whole wheat bread, alcohol, vegetables, fish, fruits and unsalted nuts ranges between 11.1 and 51.5% (Rijksinstituut voor Volksgezondheid en Milieu, 2024). Future research, using nationally representative samples, should determine whether these percentages are correct or an overestimation. Thus, in future, the HDS should therefore be thoroughly investigated in other age groups, people with different socioeconomic backgrounds, and in individuals with underlying diseases. Future studies with larger sample sizes in the general population will also enable the evaluation of possible sex and age differences.

4. Conclusion

The current analyses revealed that the HDS with and without picture examples are equally valid and reliable measures. Given the different food items across cultures, the HDS without picture examples is preferred to be used in future research. The HDS is an easy-to-use and cost-effective tool that can be implemented in for screening in clinical practice, and for research purposes (e.g., surveys or clinical trials). The HDS could also be useful in public (prevention) campaigns to create awareness among the general population about the importance of attaining a healthy diet. Future research evaluating the HDS in a nationally representative sample should determine the healthy diet percentages according to age, sex, and other demographic variables. In conclusion, the HDS is a valid and reliable scale to estimate the percentage of daily diet that is considered healthy.

Conflict of interest

The authors declare no conflict of interest. A.K. has held research grants from H2020, Nutricia-Danone, Netherlands Center of Translational Research, Lungfund, SGF/Health Holland, and NWO. J.G. is part-time employee of Nutricia Research and received research grants from Nutricia research foundation, Top Institute Pharma, Top Institute Food and Nutrition, GSK, STW, NWO, Friesland Campina, CCC, Raak-Pro, and EU. Over the past 3 years, J.V. has held research grants from Inbiose and Danone and has acted as a consultant/advisor for Eisai, KNMP, Med Solutions, Mozand, Red Bull, Sen-Jam Pharmaceutical, and Toast!. J.V. owns stock from Sen-Jam Pharmaceutical. E.O. and J.V. received travel support from Sen-Jam Pharmaceutical. The other authors have no potential conflicts of interest to disclose.

References

Alshehri, M., Kruse-Diehr, A.J., McDaniel, J.T., Partridge, J. and Null, D.B. (2023). Impact of social support on the dietary behaviors of international college students in the United States. *Journal of American College Health*, 71(8), 2436-2444. https://doi.org/10.1080/07448481.2021.1970565

Bland, J.M. and Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160. https://doi.org/10.1177/096228029900800204

Cena, H. and Calder, P.C. (2020). Defining a healthy diet: Evidence for the role of contemporary dietary patterns in health and disease. *Nutrients*, 12, 334. https://doi.org/10.3390/nu12020334

Centraal Bureau Voor de Statistiek (CBS). (2019). Opleidingsniveau. Retrieved from October 5, 2020 from website: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau [In Dutch].

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, 2nd ed. New York, USA: Lawrence Erlbaum Associates Publishers.

Crönlein, T., Langguth, B., Popp, R., Lukesch, H., Pieh, C., Hajak, G. and Geisler, P. (2013). Regensburg Insomnia Scale (RIS): a new short rating scale for the assessment of psychological symptoms and sleep in insomnia; study design: development and validation of a new short self-rating scale in a sample of 218 patients suffering from insomnia and 94 healthy controls. *Health and Quality of Life Outcomes*, 11, 65. https://doi.org/10.1186/1477-7525-11-65

Di Giosia, P., Stamerra, C.A., Giorgini, P., Jamialahamdi, T., Butler, A.E. and Sahebkar, A. (2022). The role of nutrition in inflammaging. *Ageing Research Reviews*, 77, 101596. https://doi.org/10.1016/j.arr.2022.101596

Donners, A.A.M.T., Tromp, M.D.P., Garssen, J., Roth, T. and Verster, J.C. (2015). Perceived immune status and sleep: a survey among Dutch students. *Sleep Disorders*, 2015, 721607. https://doi.org/10.1155/2015/721607

Ferrer, R.L., Burge, S.K., Palmer, R.F., Cruz, I. and RRNeT Investigators. (2016). Practical opportunities for healthy diet and physical activity: Relationship to intentions, behaviors, and body mass index. *Annals of Family Medicine*, 14(2), 109–116. https://doi.org/10.1370/afm.1886

Gans, K.M., Hixson, M.L., Eaton, C.B. and Lasater, T.M. (2000). Rate Your Plate: a dietary assessment and educational tool for blood cholesterol control. *Nutrition and Clinical Care*, 3(3), 163–169. https://doi.org/10.1046/j.1523-5408.2000.00045.x

Glasgow, R.E., Perry, J.D., Toobert, D.J. and Hollis, J.F. (1996). Brief assessments of dietary behavior in field settings. *Addictive Behaviors*, 21(2), 239–247. https://doi.org/10.1016/0306-4603(95)00056-9

Hendriksen, P.A., Tan, S., Merlo, A., van Oostrom, C.E., Bardakçi, H., Aksoy, N., Garssen, J., Bruce, G. and Verster, J.C. (2024). COVID-19 lockdown effects on sleep, immune fitness, mood, quality of life, and academic functioning: Survey data from Turkish university students. *Data* 2024, 9(2), 35. https://doi.org/10.3390/data9020035

Hoerster, K.D., Wilson, S., Nelson, K.M., Reiber, G.E. and Masheb, R.M. (2016). Diet quality is associated with mental health, social support, and neighborhood factors among Veterans. *Eating Behaviors*, 23, 168–173. https://doi.org/10.1016/j.eatbeh.2016.10.003

Koo, T.K. and Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lindström, J., Aittola, K., Pölönen, A., Hemiö, K., Ahonen, K., Karhunen, L., Männikkö, R., Siljamäki-Ojansuu, U., Tilles-Tirkkonen, T., Virtanen, E., Pihlajamäki, J. and Schwab, U. (2021). Formation and validation of the Healthy Diet Index (HDI) for evaluation of diet quality in healthcare. *International Journal of Environmental Research and Public Health*, 18, 2362. https://doi.org/10.3390/ijerph18052362

McVey, B.A., Lopez, R. and Padilla, B.I. (2021). Evidence-based approach to healthy food choices for Hispanic women. *Hispanic Health Care International*, 19(1), 17–22. https://doi.org/10.1177/1540415320921471

Mesquita, R., Janssen, D.J., Wouters, E.F., Schols, J.M., Pitta, F. and Spruit, M.A. (2013). Within-day test-retest reliability of the Timed Up and Go test in patients with advanced chronic organ failure. *Archives of Physical and Medical Rehabilitation*, 94 (11), 2131-2138. https://doi.org/10.1016/j.apmr.2013.03.024

Monteiro, C.A., Cannon, G., Lawrence, M., Costa Louzada, M.L. and Pereira Machado, P. (2019). Ultra-processed foods, diet quality, and health using the NOVA classification system. Rome, Italy: Food and Agriculture Organization of the United Nations (FAO).

Paiva, C.E., Barroso, E.M., Carneseca, E.C., de Pádua Souza, C., Dos Santos, F.T., Mendoza López, R.V. and Ribeiro Paiva, S.B. (2014). A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. *BMC Medical Research Methodology*, 14, 8. https://doi.org/10.1186/1471-2288-14-8

Paxton, A.E., Strycker, L.A., Toobert, D.J., Ammerman, A.S. and Glasgow, R.E. (2011). Starting the conversation. Performance of a brief dietary assessment and intervention tool for health professionals. *American Journal of Preventive Medicine*, 40(1), 67-71. https://doi.org/10.1016/j.amepre.2010.10.009

Rijksinstituut voor Volksgezondheid en Milieu (RIVM) (2024). Alle veranderingen. Retrieved on January 6, 2024 from website: https://www.wateetnederland.nl/resultaten/veranderingen/alle-veranderingen [In Dutch].

Shillington, K.J., Vanderloo, L.M., Burke, S.M., Ng, V., Tucker, P. and Irwin, J.D. (2021). Ontario adults' health behaviors, mental health, and overall well-being during the COVID-19 pandemic. *BMC Public Health*, 21, 1679. https://doi.org/10.1186/s12889-021-11732-6

Shim, J.-S., Oh, K. and Kim, H.C. (2014). Dietary assessment methods in epidemiologic studies. *Epidemiology and Health*, 36, e2014009. https://doi.org/10.4178/epih/e2014009

U.S. Food and Drug Administration (US FDA) (2009). Guidance for Industry. Patient-Reported Outcome Measures: use in medical product development to support labeling claims. Retrieved on February 16, 2020 from US FDA website: https://www.fda.gov/media/77832/download

Van Oostrom, E.C., Mulder, K.E.W., Verheul, M.C.E., Hendriksen, P.A., Thijssen, S., Kraneveld, A.D., Vlieg-Boerstra, B., Garssen, J. and Verster, J.C. (2022). A healthier diet is associated with greater immune fitness. *Pharma Nutrition*, 21, 100306. https://doi.org/10.1016/j.phanu.2022.100306

Verster, J.C., van de Loo, A.J.A.E., Benson, S., Scholey, A. and Stock, A.-K. (2020). The assessment of overall hangover severity. *Journal of Clinical Medicine*, 9, 786. https://doi.org/10.3390/jcm9030786

Verster, J.C., Sandalova, E., Garssen, J. and Bruce, G. (2021). The use of single-item ratings versus traditional multiple-item questionnaires to assess mood and health. *European Journal of Investigation Health Psychology and Education*, 11, 183-198. https://doi.org/10.3390/ejihpe11010015

Verster, J.C., Kraneveld, A.D. and Garssen, J. (2022). The assessment of immune fitness. *Journal of Clinical Medicine*, 12, 22. https://doi.org/10.3390/jcm12010022

Verster, J.C., Mulder, K.E.W., Hendriksen, P.A., Verheul, M.C.E., van Oostrom, E.C., Scholey, A. and Garssen, J. (2023). Test-retest reliability of single-item assessments of immune fitness, mood and quality of life. *Heliyon*, 9, e15280. https://doi.org/10.1016/j.heliyon.2023.e15280

