



Universiteit
Leiden
The Netherlands

The accuracy of portion size estimation using food images and textual descriptions of portion sizes: an evaluation study

Lucassen, D.A.; Willemsen, R.F.; Geelen, A.; Brouwer-Brolsma, E.M.; Feskens, E.J.M.

Citation

Lucassen, D. A., Willemsen, R. F., Geelen, A., Brouwer-Brolsma, E. M., & Feskens, E. J. M. (2022). The accuracy of portion size estimation using food images and textual descriptions of portion sizes: an evaluation study. *Journal Of Human Nutrition And Dietetics*, 34(6), 945-952. doi:10.1111/jhn.12878

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3276150>

Note: To cite this publication please use the final published version (if applicable).

The accuracy of portion size estimation using food images and textual descriptions of portion sizes: an evaluation study

Desiree A. Lucassen  | Romy F. Willemsen | Anouk Geelen | Elske M. Brouwer-Brolsma | Edith J.M. Feskens

Division of Human Nutrition and Health,
Wageningen University, The Netherlands

Correspondence

Desiree A. Lucassen, Division of Human
Nutrition and Health, PO Box 17, Wageningen
University, 6700 AA Wageningen, The
Netherlands.

Email: desiree.lucassen@wur.nl

Funding information

TKI Agri&Food PPS, Grant/Award Number:
AF16096

Abstract

Background: Inaccurate self-report of portion sizes is a major cause of measurement error in dietary assessment. To reduce this error, different portion size estimation aids (PSEAs) have been developed, including food images (image based, IB-PSE) and textual descriptions of portion sizes (text-based, TB-PSE). We assessed the accuracy of portion size estimation by IB-PSE and TB-PSE.

Methods: True intake of one lunch was ascertained in forty participants. Self-reported portion sizes were assessed after 2 and 24 hours by means of TB-PSE and IB-PSE, in random order. Wilcoxon's tests were used to compare mean true intakes to reported intakes. Moreover, proportions of reported portion sizes within 10% and 25% of true intake were assessed. An adapted Bland-Altman approach was used to assess agreement between true and reported portion sizes. Analyses were conducted for all foods and drinks combined and for predetermined food types.

Results: No significant differences were observed between reported portion sizes at 2 and 24 hours after lunch. Combining median relative errors of all foods items resulted in an overall 0% error rate for TB-PSE and 6% error rate for IB-PSE. Comparing reported portion sizes within 10% (31% vs. 13%) and 25% (50% vs. 35%) of the true intake showed a better performance for TB-PSE compared to IP-PSE, respectively. Bland-Altman plots indicated a higher agreement between reported and true intake for TB-PSE compared to IB-PSE.

Conclusions: Although the use of TB-PSE still results in measurement error, our results suggest a more accurate dietary intake assessment with TB-PSE than IB-PSE.

KEYWORDS

Dietary assessment, Food images, Household measures, Portion size estimation, PSEA, Standard portion sizes

INTRODUCTION

Accurate dietary assessment is essential in nutrition research. Although dietary intake is still often assessed

using paper-pencil tools, i.e. food frequency questionnaires (FFQs), food records (FRs) and 24-hour recalls (24hRs), dietary assessment techniques have advanced rapidly in recent years. The last decade numerous valuable computer-based

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Human Nutrition and Dietetics* published by John Wiley & Sons Ltd on behalf of British Dietetic Association.

and web-based tools, mostly based on 24hRs and FFQs, have been developed.¹⁻³ More recently also different smartphone applications (i.e. apps), mostly based on FRs, have been developed to collect real-time dietary intake data.^{3,4} Important benefits of these new tools include that they are assumed to lower burden on both participant and researcher compared to traditional techniques.³⁻⁶

A fundamental aspect of accurate dietary assessment is portion size estimation.⁶⁻⁸ However, assessment of portion sizes is challenging and a major cause of error in dietary assessment.^{6,9-11} Difficulties occur while reporting previously consumed foods as well as when judging displayed foods.^{4,11,12} The accuracy of portion size estimation is affected by various factors, including type of food and serving size.^{6,10,13} Generally, single-unit foods (e.g. sliced bread, fruits) are more likely to be reported correctly compared to liquids or amorphous foods (e.g. pasta, lettuce).^{4,12,14} Another issue in portion size estimation is that large portions tend to be underestimated and small portions tend to be overestimated, which is also known as the 'flat-slope phenomenon'.¹¹ In addition, foods consumed in small portions (e.g. spreads) are likely to be estimated more accurately than large portions of foods.¹³

Portion size estimation aids (PSEAs) (e.g. images, referent objects, portion size suggestions) have been suggested to result in more accurate portion sizes estimates.¹⁵⁻¹⁷ However, research indicates that these PSEAs still result in measurement error and that further optimization of PSEAs is needed¹⁷, especially with respect to PSEAs that may be implemented in web-based and smartphone-based dietary assessment tools. The most commonly used PSEAs in web-based and smartphone-based tools are portion size suggestions (i.e. standard portion sizes and household measures), food images, and free entry of weight in grams.¹ As individuals fail to recognize the metric quantities of portion sizes, estimations in grams are usually inaccurate.¹⁸ For this reason, participants tend to prefer the use of household measures rather than estimation in grams.^{17,18} Yet, inconsistent or vague descriptions of household measures may still result in measurement error, especially among individuals that are not frequently involved in meal preparation.^{18,19} Therefore, clear descriptions of the portion sizes are crucial.²⁰

To facilitate the estimation of portion sizes, several dietary assessment tools have included food images as visual aids, where individuals are requested to select the most comparable image with respect to the portion size consumed or displayed (i.e. image-based portion size assessment or IB-PSE). Previous research indicates that IB-PSE is particularly influenced by three main elements, namely perception, conceptualization and memory.¹³ Despite these elements of potential error, IB-PSE is suggested to be a useful aid to estimate portion sizes.^{14,21-24} However, there is only limited evidence on the reliability of IB-PSE in real-life situations.^{14,19} Up to now, the reliability of IB-PSE has mainly been examined by exposing participants to foods and food images simultaneously while focussing on perception and

not conceptualization and memory.²²⁻²⁴ More specifically, the majority of previous research only compared PSEAs to weighed portion sizes as a reference technique.^{12,19,21-24} To the best of our knowledge, none of the previous studies examined the accuracy of portion size estimation using a combination of textual descriptions of household measures (e.g. spoons, cups, glasses), standard portion sizes (e.g. small, medium, large) and estimation in grams (i.e. for the purpose of this study referred to as text-based portion size estimation or TB-PSE) and IB-PSE.

Therefore, the current study aimed to compare the accuracy of TB-PSE and IB-PSE. As we hypothesize that accuracy varies over different food types, accuracy of both PSEAs was examined for all foods and drinks combined and for specific food types. In addition, to gain a first insight in the effect of memory on the accuracy of the PSEAs, the portion sizes were reported after either 2 hours or 24 hours.

MATERIALS AND METHODS

Participants

Participants were recruited through a convenience sampling method using a database of research volunteers of the division of Human Nutrition and Health of Wageningen University and Research (WUR), social media accounts of the division (i.e. Facebook and Twitter), and through posters. Eligible participants were Dutch speaking, not visually impaired, not participating in another dietary intervention study, not an employee of the division, and not having any formal training in the field of nutrition. In total, 40 participants aged 20-70 years old were included in this study that was conducted during a 2-week period in February 2018. Participants were stratified by sex and age to ensure equal distribution of these characteristics and randomly assigned to two groups. Participants were informed that the study focused on different digital methods to assess food intake. The true study purpose was not disclosed until the end of the study. Written informed consent was obtained from all participants.

Overall study design

Participants were invited for one lunch at the study centre as part of the cross-over study and asked to complete two dietary questionnaires on a tablet or computer; 2 and 24 hours after lunch. The first group reported their food intake 2 hours after lunch by means of TB-PSE and 24 hours after lunch by means of IB-PSE. The second group reported their intake with the two PSEAs in the opposite order. As previous studies suggest that the potential difficulty to accurately estimate portion size depends on the type of food, we offered a variety of commonly consumed food types in the Netherlands^{7,12-14} (Table 1). Each participant was provided with pre-weighed, ad libitum amounts of the food items.

TABLE 1 Food items offered, by food type.

Offered food items
Amorphous
- Cheese
- Crunchy muesli
- Fruit salad
- Scrambled eggs
- Yogurt
Liquids
- Milk
- Orange juice
- Water
Single-units
- Bread slices
- Bread rolls
Spreads
- Jam
- Margarine

Each item was offered in a container without indication of the content. To minimize the effect of tableware on portion size estimation²⁵, the participants received a variety of tableware. After lunch, plate waste was weighed to assess true intake of each food item. Weights were taken with 'Sartorius Signum 1' calibrated weighing scales. True intake was calculated by the following formula:

$$\text{True intake (g)} = \text{Pre - weighed food item (g)} - \text{Plate waste food item (g)}$$

Portion size assessment

For the purpose of this study, a TB-PSE and IB-PSE questionnaire was developed in Qualtrics (Qualtrics, Provo, UT, USA). The question formulation and portion size estimation within the TB-PSE questionnaire were based on Compl-eat™; a self-administered web-based dietary 24hR-tool developed by WUR²⁰. Portion sizes described in Compl-eat™ are a combination of estimation in grams/millilitres, standard portion sizes and household measures, which are based on the 'Food portion sizes and coding instructions'.²⁶ The question formulation within the IB-PSE questionnaire was also based on Compl-eat™, thus ensuring that observed differences were solely due to the different PSEAs and not due to differences in question formulation. For the IB-PSE questionnaire, the portion size images from the Automated Self-Administered 24-hour dietary recall (ASA24) picture book, developed by the National Cancer Institute, Bethesda, MD²⁷, were used. This picture book contains 3 to 8 portion size images per food item. To the best of our knowledge, this is the only freely available picture book portraying food images with known amounts (g) for research purposes.²⁸ Questionnaires started with questions whether or not a type of food was consumed, which was followed by questions on the amount of food consumed by means of one of the PSEAs. An example question from each questionnaire can be found in Supplement S1.

Additional measurements

On the study day, participants completed a short questionnaire about basic characteristics (i.e. age, sex, educational level). In addition, weight and height were measured to calculate participants' BMI (kg/m^2). Participants were characterized in three educational levels (low: primary or lower education, intermediate: secondary or higher vocational education, high: college or university) and four age groups (18-28, 29-45, 46-55, 56-70 years).

Statistical analysis

Normally distributed data is displayed as means (M) and standard deviations (SDs) in case of continuous variables, or frequencies in case of categorical variables; non-normally distributed data as medians and interquartile ranges (IQRs). Significant differences between true and reported intake, and between 2 and 24 hours, were assessed for each PSEA. To allow comparison between PSEAs across different food types, relative differences were calculated. As previous research indicated that accuracy of portion size estimation varies over food types, all analyses were conducted for all foods and drinks combined and for predetermined food types individually (i.e. "all foods excluding liquids", "amorphous foods", "liquids", "single-units", "spreads"; Table 1). As there are no guidelines on the acceptable level of accuracy^{7,14,29}, the proportion of the reported intake that fell within 10% and 25% of true intake were assessed, which is in line with comparable studies in this research area.¹⁴ Proportions within 10% of true intake will be deemed acceptably accurate, whereas proportions within 25% of true intake will be used to get further insight in the levels of accuracy.³⁰ To determine agreement between reported and true intake for both PSEAs, Bland-Altman plots with 95% limits of agreement (LOA) were plotted. Usually the Bland-Altman method is applied for assessing agreement between two imperfect measures. Since true intake was assessed an adapted Bland-Altman method was used to plot the differences between reported and true intake against true intake.^{14,31} However, when true intake increased, the absolute error increased. Therefore, we plotted the log-transformed ratio of reported and true intake against log-transformed true intake. Middle line indicates the mean and the upper and lower lines indicate borders based on mean ± 1.96 SD. Since the variables were not normally distributed, Wilcoxon signed rank test was used to test within group and the Wilcoxon rank sum test was used to test for between group differences. All analyses were conducted with SAS software, version 9.4 (SAS Institute Inc., Cary, NC, USA). Statistical significance was set at $p < 0.05$.

RESULTS

A total of 40 participants took part in this study. Participants had a mean $\pm$ SD age of 46.9 ± 19.2 years (range

20.7–69.4 years), BMI 24.9 ± 3.8 kg/m², 47.5% was men and the majority of the population was highly educated (62.5%). Participant characteristics did not significantly differ between group 1 (2hR: TB; 24hR: IB) and group 2 (2hR: IB; 24hR: TB) (Table 2). Furthermore, no significant differences were observed between reported at 2 and at 24 hours after lunch, for each PSEA. Therefore, the results are only shown per PSEA and are not subdivided per time point.

Median true intake for “all foods and drinks combined” was 94 g (IQR: 128 g), while median reported intake was 75 g (IQR: 120 g) for TB-PSE and 88 g (IQR: 164 g) for IB-PSE. Comparing the true intake with the reported intake, as assessed with TB-PSE, pointed towards significant differences for “all foods excluding liquids”, “amorphous foods”, “liquids” and “spreads” (Table 3). For IB-PSE, significant differences with the true intake were observed for “all foods and drinks combined”, “liquids”, “single-units” and “spreads”. For “all foods and drinks combined” the median relative difference was 0% (IQR: 44%) as assessed by TB-PSE, and 6% (IQR: 115%) as assessed by IB-PSE (Table 3).

Significantly higher relative errors were shown for IB-PSE than for TB-PSE for “all foods and drinks combined”, “all foods excluding liquids”, “amorphous foods” and “liquids”. For “all foods and drinks combined” the proportion of reported intakes within 10% of true intake was 31% for TB-PSE and 13% for IB-PSE, the proportion within 25% of true intake was 50% for TB-PSE and 35% for IB-PSE. For TB-PSE, the lowest proportion within 10% and 25% of true intake was observed for “spreads”, whereas for IB-PSE, the lowest proportion was observed for “liquids”. The highest proportion of reported intake that fell within 10% and 25% of true intake was, for both PSEAs, observed for the food type “single-units” (Table 3).

The log-transformed Bland-Altman plot of “all foods and drinks combined” showed a higher level of agreement for TB-PSE (M: 0.04; LOA: −1.11–1.03) than for IB-PSE, as shown by more widely scattered estimates and wider limits of agreement for IB-PSE (Supplement S2). Excluding liquids did not substantially alter these findings; agreement for TB-PSE (M: −0.10; LOA: −1.22–1.00) remained higher compared

to IB-PSE (M: 0.03; LOA: −1.37–1.43). The same trend was observed for the other food types (Supplement S2). The highest level of agreement was observed for “single-units” (TB-PSE M: −0.02; LOA: −0.30–0.25 vs. IB-PSE M: −0.09; LOA: −0.84–0.66), whereas the lowest level of agreement was observed for “amorphous foods” (TB-PSE M: −0.13; LOA: −1.43–1.15 vs. IB-PSE M: 0.17; LOA: −1.38–1.71).

DISCUSSION

In this study, the reported intake and its estimation error for “all foods and drinks combined” using IB-PSE significantly differed from true intake while no statistically significant difference was observed between the reported intake and its estimation error from true intake using TB-PSE. However, as indicated by the proportion of reported intakes within 10% and 25% of true intake, being 31% and 50% using TB-PSE compared to 13% and 35% using IB-PSE, meaning that for both PSEAs only the minority of estimations lies within the acceptable range, further improvements to increase the accuracy of portion size estimation are needed.

Before discussing our findings, the strengths and limitations of our study will be discussed. First, despite the fact that participants consumed their lunch in a controlled setting, we strived to mimic a real-life situation. Specifically, in contrast to most other studies, participants could choose from a selection of food items and actually consumed the selected items.^{19,24} Furthermore, participants had the opportunity to choose between different sizes of tableware²⁵ and had ad libitum access to the foods provided.³² Moreover, all products were served in bowls, jugs and plates without indication of content. Second, as the accuracy of two PSEAs was assessed separately, accuracy of both methods could be studied independently. Moreover, due to the study's cross-over design the accuracy of both PSEAs was assessed in each participant. Third, to our knowledge, this is the first study comparing the two PSEAs, while keeping all other factors in the questionnaire identical. Finally, to avoid extra focus on portion sizes, participants were not informed on the goal of the study

TABLE 2 Characteristics of the participants.

	Total (n = 40)			Group 1 [†] (n = 20)			Group 2 [‡] (n = 20)		
	Mean	SD	%	Mean	SD	%	Mean	SD	%
Men			47.5			50.0			45.0
Age (years)	46.9	19.2		48.7	19.8		45.0	18.9	
BMI (kg/m ²)	24.9	3.8		25.9	4.1		24.0	3.3	
Educational level									
Low			0.0			0.0			0.0
Intermediate			37.5			35.0			40.0
High			62.5			65.0			60.0

Note: No significant differences were found between groups.

[†]Group 1: 2hR = TB-PSE; 24hR = IB-PSE

[‡]Group 2: 2hR = IB-PSE; 24hR = TB-PSE

TABLE 3 Median true intake, median reported intake for both PSEAs, and reported intakes within 10% and 25% of true intake for both PSEAs for all foods and per food type.

	Obs (n)	TB-PSE				IB-PSE				<10% of true (%)	<25% of true (%)	
		Median true intake (g, IQR)	Median rep. intake (g, IQR)	Median diff. (g, IQR) [†]	Median rel. diff. (%, IQR) [‡]	Median rep. intake (g, IQR)	Median diff. (g, IQR) [†]	Median rel. diff. (%, IQR) [‡]				
All food items	326	94 (128)	75 (120)	0 (25)	0 (44)	31	50	88 (164)	3 (74)***	6 (115)***	13	35
All excl. liquids	263	66 (86)	50 (101)	0 (22)****	0 (44)	32	49	61 (95)	-1 (41)	-5 (74)**	15	41
Amorphous	150	84 (124)	66 (107)	-7 (45)**	-12 (56)	12	25	104 (119)	3 (70)	4 (109)***	13	31
Liquids	63	194 (125)	220 (150)	29 (80)***	15 (39)	30	56	355 (356)	211 (373)***	118 (122)***	5	11
Single-units	59	100 (50)	100 (50)	0 (0)	0 (0)	95	95	72 (61)	-7 (25)*	-14 (18)	29	80
Spreads	54	17 (15)	15 (14)	-3 (14)*	-23 (63)	9	24	14 (12)	-3 (10)*	-22 (63)	9	30

Note: All food items, all foods and drinks combined; obs., observations; rep., reported; diff., difference; rel., relative; excl., excluding.

[†]Calculated as reported intake minus true intake. Thus, positive differences represent overestimations and negative differences represent underestimations. Significant differences between reported and true intake was assessed with a Wilcoxon signed-rank test. Significant differences are indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.0001$.

[‡]Relative differences (%) = (reported intake (g) - true intake (g)) / true intake (g) * 100. Significant differences between intake reported with TB-PSE and IB-PSE were assessed with a Wilcoxon signed-rank test. Significant differences are indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.0001$ in the IB-PSE "Mean rel. diff. column".

and did not see the weighing of the foods. A limitation of our study is that we used the ASA24 picture book in a Dutch population. The ASA24 is the only freely available photo database for research with known portion size weights. However, the ASA24 photographs are based on the 5th and 95th percentile of intake per product in the US and as such tailored for usage in the US.^{14,33,34} It is known that portion sizes in the US are larger than in the Netherlands.^{35,36} To illustrate, the glasses in the study of Donders-Engelen *et al.*²⁶ range between 100 g and 220 g whereas the glasses in ASA24 range between 177 g and 473 g. As ASA24 does not contain pictures of the smallest portion sizes consumed in the Netherlands, this may explain the overestimated intakes by IB-PSE estimates in our study (e.g. 118% for "liquids"). However, we have to note that the portion size database that currently is being used in the Netherlands dates from 2003. It is known that plate sizes have increased in the past decades³⁶, which on its turn may have led to an underestimation of TB-PSEs.

A more general limitation of the ASA24 food images is the usage of cutlery as reference, which is meant to help participants estimate the real-life size of a portion. However, as cutlery can vary in size, it might not be the best reference and as such explain the more scattered points observed in the Bland-Altman plot of IB-PSE compared to TB-PSE. Finally, in view of generalisability it needs to be mentioned that our participants were relatively old and highly educated. However, several previous studies concluded that age and education level did not affect the participants ability to estimate portion sizes.^{19,22,23,37} In addition, we only tested a limited number of food items, and as such our findings are only applicable to these tested food items.

As hypothesized, the accuracy of reported intake with both PSEAs varied between the different food types. Both PSEAs overestimated the median reported intake of "liquids" whereas the intake of "all foods excluding liquids" and "spreads" were (slightly) underestimated. In addition, for TB-PSE, the reported median intake of "amorphous foods" was underestimated, while for IB-PSE the intake was overestimated. Previous research showed both under- and overestimations of portion size estimations.^{7,14} Moreover, the accuracy of food intake estimates varied depending on the food types.^{12,13,38} Both PSEAs showed the highest estimation errors for "liquids, which is not in line with similar studies showing the highest estimation errors for "amorphous foods".^{12-14,37} In contrast to previous studies, which mostly provided liquids in containers that were identical to containers portrayed on the images, we aimed to resemble the real-life situation and therefore studied commonly-used PSEA descriptions and used glasses that did not necessarily match with the glasses on the images. As conceptualization plays a major role in the accurateness of portion size estimation¹³, it is easier to estimate portion sizes when the portion sizes are similar to the portions portrayed on the images^{23,39} or the textual descriptions.^{18,20} For instance, the description "lemonade glass" lacks detail and can easily result in misclassification. In agreement with our study, Hernandez *et al.*⁷ also studied the intake of liquids in containers that were not

identical to the containers on the images and also observed the highest estimation errors for liquids, which underlines the influence of conceptualization.

As illustrated by small errors for “single-units” and “spreads” and larg(er) errors for “amorphous foods” and “liquids” for both PSEAs, our findings clearly indicate that foods consumed in small or defined units are more accurately estimated than foods consumed in larger amounts. These findings are in line with previous studies.^{23,37,39} Generally, the accuracy for the food types “amorphous foods”, “liquids” and “single-units” was higher for TB-PSE than for IB-PSE estimates, except for “spreads” which were more accurately estimated with IB-PSE. The latter may relate to the fact that textual description of the size of spoons and spread on bread is open to interpretation, whereas a picture may provide a better impression of the portion size estimate.¹³ Moreover, the fact that we used images of spoons, instead of images of spread on bread, to estimate the amount of “spreads” consumed, may have resulted in more accurate estimates for this food type.¹² The size of the bread might influence the perception of the portion size and thereby lead to errors in estimations.²¹

We found no significant differences in accuracy between reporting after 2 hours and 24 hours for each of the PSEAs. Based on this, we concluded that memory did not influence the accuracy of portion size estimations within this time-frame. Therefore, only the combined results per PSEA were used for further analysis. However, after dividing the participants per PSEA over the two time points, the sample size per group was very small (i.e. ~20 participants) and therefore we had less power to detect significant differences. Previous research has shown that errors increase after 1-2 hours, compared to immediate estimations.²⁴ However, our first time point was after two hours and in line with our results, De Keyzer *et al.*²¹ found no increase in estimation errors after 1-2 days compared to after 4 days.²¹ To truly understand the effect of memory on accuracy of portion size estimation more research is needed with a larger sample size.

Due to lack of consensus on the minimal required level of accuracy for PSEAs no strong conclusion can be drawn on that matter. However, the accuracy of the reported intake by TB-PSE was higher than by IB-PSE for all food types except for “spreads”, which was higher with IB-PSE. Overall, TB-PSE provided more accurate portion size estimations than IB-PSE. As discussed, these findings are different from previous studies.^{14,21-24} However, in contrast to these studies we incorporated all elements that influence IB-PSE (i.e. perception, conceptualization, memory), instead of focusing on one or two of these elements²²⁻²⁴, in an attempt to mimic a real-life situation. Therefore, our findings in combination with previous studies may indicate that IB-PSE is a useful PSEA, but only when judging displayed foods and not for retrospective portion size estimation.

TB-PSE and IB-PSE were selected due to their applicability for implementation in web-based and smartphone-based dietary assessment tools. However, there are other PSEAs which would be applicable for implementation in web-based or smartphone-based dietary assessment tools (e.g. remote food photography method, body-worn

monitors).^{8,40} These innovative tools also have a range of drawbacks, for instance, it is known that they are unable to detect all aspects of the food consumed (e.g. no difference detected between spinach vs. spinach a la crème).⁴¹ Furthermore, individuals might feel uncomfortable wearing the device, especially long-term, and it is difficult to guarantee the privacy of bystanders.⁴⁰ Moreover, even though these devices have been proven to be up to 90% accurate⁴⁰, such devices are expensive and therefore not suited for large-scale studies. Selecting a PSE-tool needs to be considered carefully while taking into account study design, methods and target group.⁸ Therefore, even though there are new, more innovative PSE-tools being developed, it is still valuable to further improve both TB-PSE and IB-PSE. These PSEAs are easy to implement in web-based and smartphone-based tools, relatively inexpensive, well-known and therefore easy to use with limited training.

To conclude, in our study TB-PSE is shown to be more accurate than IB-PSE. Country-specific pictures with a clear reference are needed to improve the accuracy of IB-PSE. Next to this, we can conclude that TB-PSE seems to be an accurate PSEA for “single-units”, as 95% of the reported intake fell within 10% of true intake. However, for the other food types, only 32% or less of the reported intakes fell within 10% of truth. Therefore, in line with Bucher *et al.*⁴², we conclude that the accuracy of portion size estimations with TB-PSE needs to be improved further and therefore standardized terminology is needed to avoid ambiguity with regard to textual descriptions of portion sizes. Finally, the use of a combination of PSEAs might be valuable to increase accuracy of portion size estimation.

ETHICS STATEMENT

According to the Central Committee on Research involving Human Subjects (CCMO), this type of study did not require approval from an ethics committee in the Netherlands.

ACKNOWLEDGMENTS

The authors want to thank Els Siebelink for her assistance in the design and data collection of this study. Furthermore, the authors would like to thank Hendriek Boshuizen and João Paulo for their advice with regard to the data analysis. The present study was executed by Wageningen University and Research and funded by the Ministry of Agriculture, Nature and Food Quality and industry, in the context of TKI Agri&Food PPS – project Smart Food Intake (AF16096). The funders had no role in the design, analysis and interpretation of the data or in writing this article.

CONFLICT OF INTEREST

The authors had no financial or personal conflicts of interest to declare.

AUTHOR CONTRIBUTION

The authors contributions are as follows: D.A.L. contributed to the study design, data collection, interpretations of the findings, data analysis and drafted the manuscript; R.F.W.

contributed to the study design, data collection, data analysis, interpretation of the findings and revised earlier versions of the manuscript; A.G. contributed to the study design, interpretations of the findings and revised earlier versions of the manuscript; E.B.B. and E.J.M.F. contributed to the interpretations of the findings and revised earlier versions of the manuscript. All authors read and approved the final version of the manuscript.

TRANSPARENCY DECLARATION

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported. The reporting of this work is compliant with STROBE guidelines. The lead author affirms that no important aspects of the study have been omitted and that any discrepancies from the study as planned have been explained. The study has not been registered in any trials registry.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/jhn.12878>.

ORCID

Desiree A. Lucassen  <https://orcid.org/0000-0001-9769-6566>

REFERENCES

- Timon CM, van den Barg R, Blain RJ, Kehoe L, Evans K, Walton J, et al. A review of the design and validation of web- and computer-based 24-h dietary recall tools. *Nutr Res Rev*. 2016;29:268–80.
- Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L, et al. A review of the use of information and communication technologies for dietary assessment. *Br J Nutr*. 2009;101(Suppl 2):S102–112.
- Shim JS, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. *Epidemiol Health*. 2014;36:e2014009.
- Thompson FE, Subar AF. Dietary Assessment Methodology. In *Nutrition in the Prevention and Treatment of Disease*. Elsevier Inc.; 2017. p. 5–48.
- Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Research*. 2017;6:926.
- Walton J. Dietary Assessment Methodology for Nutritional Assessment. *Topics in Clinical Nutrition*. 2015;30:33–46.
- Hernández T, Wilder L, Kuehn D, Rubotzky K, Moser-Veillon P, Godwin S, et al. Portion size estimation and expectation of accuracy. *J Food Compos Anal*. 2006;19:S14–S21.
- Sharma V, Chadha R. Effectiveness of food portion size estimation aids for diet assessment: a systematic review. *Int J Food Sci Nutr*. 2017;2:106–12.
- Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhauser ML, et al. Addressing current criticism regarding the value of self-report dietary data. *J Nutr*. 2015;145:2639–45.
- Vuckovic N, Ritenbaugh C, Taren DL, Tobar M, et al. A qualitative study of participants' experiences with dietary assessment. *J Am Diet Assoc*. 2000;100:1023–8.
- Young LR, Nestle MS. Portion sizes in dietary assessment: issues and policy implications. *Nutr Rev*. 1995;53:149–58.
- Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *J Am Diet Assoc*. 2010;110:55–64.
- Nelson M, Atkinson M, Darbyshire S. Food photography. I: the perception of food portion size from photographs. *Br J Nutr*. 1994;72:649–63.
- Kirkpatrick SI, Potischman N, Dodd KW, Douglass D, Zimmerman TP, Kahle LL, et al. The use of digital images in 24-hour recalls may lead to less misestimation of portion size compared with traditional interviewer-administered recalls. *J Nutr*. 2016;146:2567–73.
- Byrd-Bredbenner C, Schwartz J. The effect of practical portion size measurement aids on the accuracy of portion size estimates made by young adults. *J Hum Nutr Diet*. 2004;17:351–7.
- Faggiano F, Vineis P, Cravanzola D, Pisani P, Kompero G, Riboli E, et al. Validation of a method for the estimation of food portion size. *Epidemiology*. 1992;3:379–82.
- Faulkner GP, Livingstone MBE, Pourshahidi LK, Spence M, Dean M, O'Brien S, et al. An evaluation of portion size estimation aids: precision, ease of use and likelihood of future use. *Public Health Nutr*. 2016;19:2377–87.
- Spence M, Livingstone M, Hollywood LE, Gibney ER, O'Brien SA, Pourshahidi L, et al. A qualitative study of psychological, social and behavioral barriers to appropriate food portion size control. *Int J Behav Nutr Phys Act*. 2013;10:92.
- Biltoft-Jensen A, Holmgaard Nielsen T, Hess Ygil K, Christensen T, Fagt S. Accuracy of food photographs for quantifying food servings in a lunch meal setting among Danish children and adults. *J Hum Nutr Diet*. 2018;31:131–40.
- Meijboom S, van Houts-Streppel MT, Perenboom C, Siebelink E, van de Wiel AM, Geelen A, et al. Evaluation of dietary intake assessed by the Dutch self-administered web-based dietary 24-h recall tool (Compl-eat™) against interviewer-administered telephone-based 24-h recalls. *J Nutr Sci*. 2017;6:e49.
- De Keyser W, Huybrechts I, De Maeyer M, Ocké M, Slimani N, van 't Veer P, et al. Food photographs in nutritional surveillance: errors in portion size estimation using drawings of bread and photographs of margarine and beverages consumption. *Br J Nutr*. 2011;105:1073–83.
- Naska A, Valanou E, Peppas E, Katsoulis M, Barbouni A, Trichopoulou A, et al. Evaluation of a digital food photography atlas used as portion size measurement aid in dietary surveys in Greece. *Public Health Nutr*. 2016;19:2369–76.
- Ovaskainen M-L, Paturi M, Reinivuo H, Hannila M-L, Sinkko H, Lehtisalo J, et al. Accuracy in the estimation of food servings against the portions in food photographs. *Eur J Clin Nutr*. 2008;62:674–81.
- Szenczi-Cseh J, Horvath Z, Ambrus A. Validation of a food quantification picture book and portion sizes estimation applying perception and memory methods. *Int J Food Sci Nutr*. 2017;68:960–72.
- Van Ittersum K, Wansink B. Plate size and color suggestibility: the Delboeuf Illusion's bias on serving and eating behavior. *J Consum Res*. 2011;39:215–28.
- Donders-Engelen MR, Van der Heijden LJM, Hulshof KFAM. (2003) *Maten, Gewichten en Codenummers 2003. Food Portion Sizes and Coding Instructions*. Division of Human Nutrition and TNO Nutrition.
- Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. *J Acad Nutr Diet*. 2012;112:1134–7.
- ASA24. ASA24: Automated Self-Administered 24-hour dietary assessment tool; 2018. <https://epi.grants.cancer.gov/asa24/respondent/citations.html> (accessed 30-04-2018).
- European Food Safety Authority. Use of the EFSA Comprehensive European Food Consumption Database in Exposure Assessment. *EFSA Journal*; 2011;9:2097.
- Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. *Nutr J*. 2015;14:40.
- Krouwer JS. Why Bland-Altman plots should use X, not (Y+ X)/2 when X is a reference method. *Stat Med*. 2008;27:778–80.
- Clancy AK, Woods K, McMahon A, Probst Y. Food composition database format and structure: a user focused approach. *PLoS One*. 2015;10:e0142137.
- Kirkpatrick SI, Subar AF, Douglass D, Zimmerman TP, Thompson FE, Kahle LL, et al. Performance of the Automated Self-Administered 24-hour Recall relative to a measure of true intakes

- and to an interviewer-administered 24-h recall. *Am J Clin Nutr*. 2014;100:233–40.
34. Kirkpatrick SI, Thompson FE, Subar AF. Validity of the National Cancer Institute's Automated Self-Administered 24-hour Recall (ASA24): results of a feeding study. *FASEB J*. 2013;27:230.234.
 35. Rozin P, Kabnick K, Pete E, Fischler C, Shields C. The ecology of eating: smaller portion sizes in France Than in the United States help explain the French paradox. *Psychol Sci*. 2003;14:450–4.
 36. Steenhuis IH, Leeuwis FH, Vermeer WM. Small, medium, large or supersize: trends in food portion sizes in The Netherlands. *Public Health Nutr*. 2010;13:852–7.
 37. Bucher T, Weltert M, Rollo ME, Smith SP, Jia W, Collins CE, et al. The international food unit: a new measurement aid that can improve portion size estimation. *Int J Behav Nutr Phys Act*. 2017;14:124.
 38. Robson PJ, Livingstone MBE. An evaluation of food photographs as a tool for quantifying food and nutrient intakes. *Public Health Nutr*. 2000;3:183–92.
 39. Vereecken CA, Covents M, Matthys C, Maes L, et al. Young adolescents' nutrition assessment on computer (YANA-C). *Eur J Clin Nutr*. 2005;59:658–67.
 40. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary intake and physical activity assessment: current tools, techniques, and technologies for use in adult populations. *Am J Prev Med*. 2018;55:e93–e104.
 41. Fengqing Zhu, Bosch M, Insoo Woo, SungYe Kim, Boushey CJ, Ebert DS, et al. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE J Sel Top Signal Process*. 2010;4:756–66.
 42. Bucher T, Rollo ME, Smith SP, Dean M, Brown H, Sun M, et al. Position paper on the need for portion-size education and a standardised unit of measurement. *Health Promot J Austr*. 2017;28:260–3.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lucassen DA, Willemsen RF, Geelen A, Brouwer-Brolsma EM, Feskens EJ. The accuracy of portion size estimation using food images and textual descriptions of portion sizes: an evaluation study. *J Hum Nutr Diet*. 2021;34:945–952. <https://doi.org/10.1111/jhn.12878>