Empirical: Single or Multiple Studies

Check for updates

Using Food Frequency Questionnaires to Measure Traits: A Case Study of Human Consumption of Animals and Animal Products

Adam Feltz [1,2] , Jacob Caton [3], Zac Cogley [4] , Mylan Engel, Jr. [5] , Silke Feltz [1],

Ramona Ilea [6] , L. Syd M. Johnson [7] , Tom Offer-Westort [1]

[1] Department of Psychology, University of Oklahoma, Norman, OK, USA. [2] Center for Applied Social Research, University of Oklahoma, Norman, OK, USA. [3] Department of English and Philosophy, Arkansas State University, Jonesboro, AR, USA. [4] Balto Software, Saint Louis, MO, USA. [5] Northern Illinois University, DeKalb, IL, USA. [6] Pacific University, Forest Grove, OR, USA. [7] Center for Bioethics and Humanities, Upstate Medical University, Syracuse, NY, USA.

Corresponding Author: Adam Feltz, Department of Psychology, University of Oklahoma, 455 W. Lindsey St., Dale Hall Tower, Room 705, Norman, OK 73019, USA. E-mail: afeltz@ou.edu

Supplementary Materials: Materials, Preregistration [see Index of Supplementary Materials]

Abstract

Measuring human consumption of animals and animal products (HCAAP) is challenging but often important for researchers and animal rights advocates. We contribute to measuring HCAAP by conceptualizing that consumption as a trait. In 3 studies, we analyzed responses from traditional Food Frequency Questionnaires and created two measures of HCAAP traits based on 24-hour and 3-month self-reports. Studies 1 (N = 249) and 2 (N = 265) evaluated the item-level properties of 24-hour and 3-month self-reports, eliminating items that were not likely to provide much information about the underlying trait of HCAAP. Study 3 (N = 252) provided evidence that the two measures were predicted by knowledge of animals as food, meat-eating rationalizations, numeracy, sex, and political orientation. These results suggest that the two instruments could be used to measure HCAAP as a trait. We offer suggestions as to when using the two instruments may be beneficial.

Keywords

Non-Technical Summary

Background

Predicting and reducing human consumption of animals and animal products are among the central aims of researchers and activists. Among the ways that researchers have measured human consumption of animals and animal products is with food frequency questionnaires. Food frequency questionnaires involve asking participants how much of certain kinds of food (e.g., beef) one consumes over some period (e.g., over the previous 24 hours). Researchers have identified several shortcomings with food frequencies questionnaires. These shortcomings include, but are not limited to, many items having low frequency of consumption (e.g., liver), large variation in tendencies to consume some animal products, and the general shortcomings associated with having 1–item measures of variables (e.g., estimating reliability of the responses). These shortcomings can make it difficult to predict and detect changes in human consumption of animals and animal products.

Why was this study done?

Our studies were conducted to help provide one potential solution to the problems associated with traditional food frequency questionnaires. Instead of conceptualizing and measuring human consumption of animals and animal products with single items, we offer an alternative conceptualization of human consumption of animals and animal products as a trait. One way to conceptualize traits is that they are dispositions of people to think, feel, or behave that are stable, internally caused, are frequent displayed, and occur in many different situations. On this conceptualization, human consumption of animals and animal products can be thought of as a trait because the consumption tends to persist over time, are decided by individuals, happen frequently, and happen in many different situations.

What did the researchers do and find?

We used items from two different food frequency questionnaires to find a way to measure human consumption of animals and animal products as a trait. One centered on retrospective self-reports of the amount of consumption over the past 3 months and the other centered on retrospective self-reports of consumption over the past 24 hours. Each instrument started with 25 animals and animal products. We then engaged in a process of item-reduction using Item Response Theory to identify items that would provide the most information about a range of human consumption of animals and animal products. Items were selected that estimated a range of consumption behaviors along with being able to discriminate people at different consumption levels. In three studies, these methods identified 6 items for the 3-month measure and 13 items for the 24-hour measure that provided reasonable estimates of human consumption of animals and animal products traits. The means for each of these measures were related to each other and were related to rationalizations for animal consumption, knowledge of animals used as food, and a measure of general decision-making skill.

What do these findings mean?

These results suggest that there might be value in conceptualizing human consumption of animals and animal products as a trait. If the two measures offer a useful way of measuring human consumption of animals and animal products, then using those instruments should provide a new way for researchers to predict and change human consumption of animals and animal products. Using the instruments in these ways could help overcome difficulties associated with only using single items in traditional food frequency questionnaires.

Many researchers and activists have been interested in measuring people's propensity to consume animal products. Two reasons why researchers are interested in measuring consumption involve predicting human consumption of animals and animal products (HCAAP) and estimating the effectiveness of interventions to change consumption. However, limitations in some measurements of HCAAP have sometimes made estimating these relations and changes difficult. To overcome some of these limitations, we adopted a novel view that HCAAP behaviors can be conceived of as a trait. In three studies, we provide evidence for the utility of conceiving HCAAP as a trait and for ways to measure that trait. We used responses from traditional Food Frequency Questionnaires to create one measure of HCAAP based on 24-hour retrospective self-reports and another measure based on 3-month retrospective self-reports. We close by offering suggestions about future directions and when and for whom the new measures might be useful.

Measuring Human Consumption of Animals and Animal Products

There are many potential ways to measure HCAAP. What we mean by animal products are foods produced by animals like eggs and dairy. Probably the most used methods to measure HCAAP involve retrospective self-reports to Food Frequency Questionnaires (FFQs). FFQs are characterized by asking participants to report their eating behaviors over a period of time. The nature of the self-reports can include asking weights of food consumed, volume of foods consumed, servings of food consumed, or number of instances when food was consumed, just to name a few. The periods of time participants are asked to report about can also vary. Some common intervals are 24 hours, 7 days, and 3 months (for a review, see Cade et al., 2002).

There are many challenges associated with using FFQs to measure HCAAP. We identify two of these challenges here. First, there is often wide variation of responses with specific FFQ items (e.g., beef or chicken) (see NIH "Difficulties Posed By Intra-Individual Variation" for helpful examples). The wide variation in responses can make it difficult to predict consumption or to measure change in animal consumption. To illustrate, according to the NIH NHANES estimates, the mean poultry consumption in the United States over 24 hours is 1.30 servings with a standard deviation of 3.06 (see NIH "Usual Dietary Intakes"). Assuming 3.06 is the pooled standard

deviation and the reduction in HCAAP is 1.30 servings per day on average, the maximum reduction in poultry consumption cannot exceed .42 standard deviations (1.30/3.06 = 0.42). To reliably detect a reduction of .42 standard deviations (with power = .80) in a between subjects design, one would need 222 total participants if poultry consumption is completely eliminated in one group—an unlikely outcome. A more reasonable reduction would be to reduce poultry consumption by half between a control and experimental group (d = .21). In that case, one would need a total of 877 participants overall to achieve acceptable power (see Supplementary Materials for a power curve for the non-linear relation between effect size and sample size needed). For many researchers, those sample sizes are practically out of reach. Consequently, researchers may have an effective intervention but may not have access to sufficiently powered studies to be able to detect conventionally significant effects in experiments if they use HCAAP measures with large variances.

Second, potential problems have been identified with using responses from single items in FFQs. One problem with single item measures is that many food items are seldom consumed (e.g., liver) (Kipnis et al., 2009). The lack of consumption results in not being able to easily detect changes in those HCAAP behaviors since people already rarely consume those products (Cade et al., 2002). A second issue with one-item measures is that they tend to risk being more unreliable than multi-item measures. Of course, there are important exceptions to this general tendency (see, for example, Gardner et al., 1998). That having been said, statistically one-item measures cannot demonstrate some of the internal properties (e.g., internal reliability) that multi-item measures can (Loo, 2002). Corrections are available to help with some of the known biases and issues with single item measures (Freedman et al., 2014). However, many of these corrections are resource intensive (e.g., test–retest, large samples), and consequently are still practically out of reach for many researchers and activists. For these reasons and others, using one-item measures of HCAAP (e.g., beef) can be problematic and result in an inability to detect conventionally significant relations.

Rather than estimating HCAAP from single items that can be substantially skewed or have large variances, one may find a solution to those problems by trying to estimate general patterns of HCAAP. In some instances, inter-related behavioral tendencies to consume animals and animal products can be characterized as a *trait*. There are multiple ways to conceive of and measure traits (Fridhandler, 1986). We follow Chaplin et al. (1988) in defining a trait (as opposed to things such as states) along the following dimensions: duration and stability (how enduring the tendency is), locus of causation (the tendency should be internally and not situationally caused), frequency (how often the tendency manifests itself), and situational scope (the number of different situations the tendency occurs in). Our initial conceptualization of HCAAP traits was reflective. That is, there is some common, unobserved variable that causes responses to the indicator items (Myszkowski et al., 2019).

PsychOpen GOLD

Perhaps the best-known kinds of traits are personality traits. Personality traits are stable dispositions for people to feel and act in certain ways. They are temporally stable, display themselves in a variety of situations, are caused by factors internal to individuals, and differentiate people from one another. Traits need not determine behaviors, but the traits should increase the probability that one thinks or behaves in trait-consistent ways across situations (see Haslam et al., 2017). Of note, there are no well accepted criteria to determine when something qualifies as a trait on these dimensions. For that reason, we also follow Chaplin et al. (1988) in understanding traits as a prototype concept where there are some clear cases of traits, some clear cases of non-traits, and some cases that are unclear or borderline. We also leave it open whether traits must be psychological or can also be behavioral (Buss, 1985). Finally, while traits tend to be enduring, that does not mean that one cannot change traits either over time or through interventions (see, for a review, Roberts et al., 2017). On this conception, if HCAAP can be characterized as a trait, then the tendency to consume animals should persist for a long period of time, should be caused by factors internal to the person, should happen frequently and in many different situations.

Traits are often measured by using responses to multiple questions or prompts. For example, one common measure of the HEXACO model of personality (Honest-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, Openness to experience) uses 100 items to measure those 6 personality traits (Lee & Ashton, 2018). Each of the individual traits has a set of items devoted to measuring the specific trait. In the 100–item measure, extraversion is measured by responses to 16 items (e.g., "I enjoy having lots of people around to talk with."). These 16 items have gone through a thorough vetting process to provide evidence that those items likely measure the underlying (i.e., latent) trait of extraversion. Only using 1 or 2 items from that 16-item measure would, without further validation, likely reduce the quality of the measurement of extraversion. With further studies, we could have evidence that a different or smaller set of items could measure extraversion (e.g., there is a 60 item measure of the HEXACO, among numerous other validated personality measures; Ashton & Lee, 2009). But what is common to all these measures is that they have gone through a vetting process to make sure that the responses to the set of items provides information about the target underlying trait.

Given this conception of traits, we decided to measure HCAAP traits with retrospective self-reports of how frequently one consumes animals and animal products. We used items from traditional FFQs to elicit these self-reports. We tested these items using factor analytic techniques to find the items that would likely provide the most information about the underlying trait of HCAAP. To our knowledge, ours is the first attempt to use responses to FFQs to measure HCAAP traits rather than using those FFQ items to estimate specific HCAAP amounts. In 3 studies, we provide evidence that using retrospective self-reports of HCAAP can estimate the underlying trait-like tendency of HCAAP.

PsychOpen GOLD

Study 1

Study 1 tested an initial battery of items taken from common FFQs. Participants retro-spectively self-reported consumption behaviors over both 24 hours and 3 months. We gathered responses from 24-hour and 3-month FFQs in parallel mainly for two reasons. First, some evidence suggests that 24-hour measures are better than 3-month measures for intervention research (Thompson et al., 2015). Second, we used two FFQs because we wanted to compare the results of the two instruments. If both instruments measure the same HCAAP trait, then the two instruments should be correlated. While the 24-hour window might be too small to measure traits (as opposed to, for example states), there is some reason to think that the 24-hour measure might measure trait-like tendencies (see, for example, evidence that some measures of trait and short-term state anxiety measure both trait-like and state-like tendencies; Lance et al., 2021). The primary goal of Study 1 was to test item–level properties of the 24-hour and 3-month items to select the items that had desirable measurement properties. Pre-registration of the study and all study items are available at Supplementary Materials. Study 1 was the only pre-registered study of the three studies. Studies 2 and 3 were based on the results of the previous study. Data are available on request.

Method

Participants

Two hundred fifty-eight participants were recruited from Amazon's Mechanical Turk. For tasks involving responses to questionnaires, Amazon's Mechanical Turk is an accept-able and often better way of recruiting subjects when compared to traditional methods (e.g., universities' subjects pools; Heen, Lieberman, & Miethe, 2014). Nine participants were excluded for incorrectly answering a comprehension question ("How many times have you consumed any food over the past three months?" The inattentive answer was "never"). Fifty-six percent identified as female ($N = 139$) and the rest identified as male with a mean age of 39.53, $SD = 12.64$.

Materials

Participants were asked how frequently they ate 25 different foods derived from ani-mals over 24 hours and over 3 months. These items were selected because they were recommended by Animal Charity Evaluators and EPIC Norfolk. For the 3-month items, participants were given the following instructions: "How often, in the past 3 months, did you eat the following?" Participants could respond: Never, Less than 1 time per week, 1–3 times per week, 4–6 times per week, or 1 or more times per day (coded 1–5). For the 24-hour items, participants were instructed "In the past day, how many times did you consume the following food and drinks?". Participants could respond from 0 to 10+ times

(coded 1–11). After completing each set of food items, basic demographic information was gathered. The food items and descriptive statistics are listed in Table 1.

Results and Discussion

Planned analyses proceeded by analyzing the item-level properties for each of the 3-month and 24-hour food items separately (see means, standard deviations, and skewness in Table 1). We first analyzed responses to the 3-month items. To explore the factor structure of these items, an exploratory factor analysis was conducted on the items using parallel analysis with minimal residual extraction and oblimin rotation. Two factors were identified with eigenvalues greater than 1 (see Supplementary Materials). The two factors were largely identified by the frequency of consumption. Sixteen items were infrequently consumed and loaded on the same factor.

The same analyses were conducted on the 24-hour items. A visual inspection of the histograms suggested that almost all the items were seldomly consumed and had substantial positive skew. An exploratory factor analysis using parallel analysis with minimal residual extraction and oblimin rotation revealed one factor with an eigenvalue greater than 1 (see Supplementary Materials). While the exploratory factor analysis only identified one factor, this was mostly likely attributable to the substantial skew and infrequent consumption of most items (skewness for all items > 1.65).

Study 1 suggested that some of the items used to measure HCAAP were problematic, largely because many food items were rarely consumed. For the 3-month measure, 16 items were not consumed frequently enough to provide much information about HCAAP traits (*item minimum mean* = 1.31, *SD* = 0.75, *item maximum mean* = 2.19, *SD* = 0.98, *factor mean* = 1.74, *SD* = 0.61, or less than less than 1 time a week on average): turkey, other meat, corned beef, sausages, savory pies, liver, hot dogs, processed meat, tuna, sherbet, cottage or ricotta cheese, sour cream, ice cream, chicken wings, cream cheese, and chicken nuggets. Those 16 items were eliminated in subsequent studies. For the 24-hour measure, discriminating between good and bad items was more difficult because all the items displayed substantial positive skew and loaded on the same factor. Some insights about which items are likely to provide information about the underlying trait can be had by looking at items not consumed frequently. The following items had consumption values less than 'one time or less' in the previous 24 hours (*item minimum mean* = 1.52, *SD* = 1.69, *item maximum mean* = 2.01, *SD* = 2.19, *13-item average* = 1.72, *SD* = 1.79): other meat, corned beef, savory pies, liver, hot dogs, tuna, sherbet, chicken nuggets, cottage cheese, sour cream, ice cream, cream cheese, and chicken wings. Using a value of 2.01 as a cutoff for low consumption, those 13 items were excluded from subsequent studies.

Table 1

Overall Means (M), Standard Deviations (SD), and Skewness for the Items in the Food Frequency Questionnaires in Studies 1–3

Item	Study	3-Month			24-Hour		
		M	SD	Skew	M	SD	Skew
Dairy (cheese, milk, yogurt, etc.)	1	3.74	1.00	-0.42	3.37	2.26	1.69
	2	3.81	1.03	-0.48	2.85	1.22	0.70
	3	—	—	—	3.11	1.43	0.35
Chicken (fried chicken, in soup, grilled chicken, etc.)	1	3.06	0.84	-0.24	2.41	2.23	2.25
	2	3.16	0.98	-0.35	2.22	1.32	1.19
	3	3.16	0.85	-0.27	2.23	1.33	1.13
Turkey (turkey dinner, turkey sandwich, in soup, etc.)	1	2.11	0.97	0.82	1.83	1.99	2.97
	2	—	—	—	1.93	1.51	1.49
	3	—	—	—	1.79	1.32	1.72
Fish and Seafood (tuna, shrimp, crab, etc.)	1	2.30	0.93	0.29	1.92	2.07	2.76
	2	2.53	0.98	0.29	1.92	1.45	1.52
	3	2.46	0.89	0.55	1.87	1.30	1.70
Pork (ham, pork chops, ribs, etc.)	1	2.21	0.93	0.53	1.94	2.07	2.83
	2	2.57	0.96	0.17	1.95	1.43	1.47
	3	2.38	0.92	0.34	1.81	1.23	1.63
Beef (steak, meatballs, in tacos, etc.)	1	2.62	0.93	-0.24	2.17	2.19	2.40
	2	2.87	0.96	-0.19	2.14	1.42	1.18
	3	2.78	0.98	-0.14	2.00	1.29	1.34
Other meat (duck, lamb, venison, etc.)	1	1.47	0.82	2.10	1.64	1.92	3.22
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Eggs (omelet, in salad, in baked goods, etc.)	1	3.07	1.03	-0.24	2.40	2.18	2.19
	2	3.22	0.98	-0.08	2.32	1.43	1.08
	3	—	—	—	2.36	1.37	1.00
Bacon	1	2.03	1.00	0.63	1.92	2.15	2.76
	2	2.49	1.05	0.51	1.92	1.40	1.50
	3	2.39	1.02	0.66	1.87	1.36	1.71
Corned Beef	1	1.43	0.83	2.16	1.59	1.88	3.32
	2	—	—	—	—	—	—

Item	Study	3-Month			24-Hour		
		M	SD	Skew	M	SD	Skew
	3	—	—	—	—	—	—
Sausages	1	1.90	0.94	1.05	1.79	1.97	2.88
	2	—	—	—	1.92	1.48	1.44
	3	—	—	—	1.84	1.33	1.62
Savory pies (e.g., meat pie, pork pie, pasties, steak & kidney pies, sausage rolls)	1	1.45	0.86	2.34	1.57	1.81	3.46
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Liver, liver pate, liver sausage	1	1.31	0.75	2.79	1.53	1.69	3.32
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Hot dogs	1	1.83	0.87	0.97	1.72	1.93	3.07
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Processed meats (e.g., salami, bologna, etc.)	1	1.93	1.02	1.00	1.78	1.87	2.94
	2	—	—	—	1.92	1.41	1.56
	3	—	—	—	1.81	1.30	1.61
Canned tuna fish	1	1.85	0.95	0.82	1.70	1.93	3.16
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Sherbet	1	1.35	0.80	2.74	1.54	1.73	3.53
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Cottage or ricotta cheese	1	1.68	0.94	1.46	1.87	2.17	2.69
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Sour cream	1	1.90	0.97	1.00	1.79	2.02	2.91
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Ice cream	1	2.19	0.98	0.53	2.01	1.19	2.61
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Chicken wings	1	1.85	0.92	0.84	1.79	2.12	2.82
	2	—	—	—	—	—	—

PsychOpen GOLD

Item	Study	3-Month			24-Hour		
		M	SD	Skew	M	SD	Skew
	3	—	—	—	—	—	—
Cream cheese	1	1.83	0.92	1.06	1.85	2.23	2.82
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Chicken nuggets	1	1.83	0.96	1.31	1.74	2.09	3.07
	2	—	—	—	—	—	—
	3	—	—	—	—	—	—
Hamburgers	1	2.19	0.96	0.54	1.94	2.13	2.69
	2	2.51	1.05	0.51	1.96	1.46	1.39
	3	2.48	0.96	0.48	1.84	1.33	1.54
Meat (any type of meat, including beef, pork, chicken, turkey, fish shellfish, or other meats)	1	3.30	1.25	-0.47	2.99	2.50	1.65
	2	2.68	1.06	-0.76	2.76	1.39	0.65
	3	—	—	—	2.78	1.49	0.62

Note. '3-Month' refers to the items in the 3-month Food Frequency Questionnaire and '24-Hour' refers to the items in the 24-hour Food Frequency Questionnaire.

PsychOpen GOLD

Study 2

Study 2 was designed to retest the items retained from Study 1. Study 2 was also designed to help provide convergent and criterion validity for the two FFQs.

Method

Participants

Two hundred sixty-five participants were recruited from Amazon's Mechanical Turk. Forty-six percent (N = 121) of participants identified as female and the rest identified as male. The mean age was 36.55, SD = 11.40.

Materials

Participants received modified versions of the FFQs used in Study 1. For the 3-month FFQ, participants were given the same instructions used in Study 1 and were asked to rate how frequently they consumed the following food items: dairy, chicken, fish, pork, beef, eggs, bacon, any kind of meat, and hamburgers. For the 24-hour FFQ, participants received the same instructions with slightly modified response options. In effort to reduce the positive skew associated with all the items in Study 1, frequency of consumption response options were changed to 0–5+ times in the previous 24 hours (coded 1–6). For the modified 24-hour FFQ, participants indicated how frequently they consumed following items: dairy, chicken, turkey, fish, pork, beef, eggs, bacon, sausages, processed meats, hamburgers, any kind of meat.

After completing the FFQs, participants received the following instruments, in order. These instruments were selected because they have been associated with HCAAP. If the two FFQs measure HCAAP tendencies, then the instruments should be correlated with each of these measures.

The Knowledge of Animals as Food Scale — The Knowledge of Animals as Food Scale (S. Feltz & Feltz, 2019b) is a 9–item, true/false objective measure of what people know about animals used as food. Previous research has suggested that this measure is negatively related to HCAAP (A. Feltz & Feltz, 2021; S. Feltz & Feltz, 2019a, 2019b) Correct answers were coded as 1 and incorrect answers as 0. The total number of correct answers was used in all analyses.

Meat-Eating Rationalizations — Meat-eating rationalizations (Piazza et al., 2015) were measured using the 4Ns instrument. The 4Ns instrument is a 16–item, Likert scale (1 = completely disagree, 7 = completely agree) measure of attitudes concerning how Natural, Normal, Necessary, and Nice eating animals and animal products is. Higher scores on the 4Ns instrument have been associated with greater HCAAP. An average of responses to the 16–items was used in all analyses.

The Berlin Numeracy Test — The Berlin Numeracy Test (Cokely et al., 2012) is a 7–item, fill-in-the-blank measure of numeracy. Numeracy refers to one's ability to use and understand statistical information. Higher performance on the Berlin Numeracy Test has been associated with lower consumption of animals and animal products (S. Feltz & Feltz, 2019b). Correct answers were coded as 1, incorrect answers as 0, and a total of correct answers was used in all analyses.

Basic demographic information was gathered including a measure of political orientation (1 = very liberal, 7 = very conservative). Previous research suggests that being more politically conservative and identifying as male are associated with consuming more animal products (Lusk & Norwood, 2016).

Results and Discussion

Analyses proceeded in two steps. First, we analyzed the item-level properties of the items in each of the 3-month and 24-hour FFQs using a graded response Item Response Theory model (Baker, 2017; Rizopoulos, 2006). Second, we estimated correlations between the FFQs and the other instruments gathered in Study 2.

One important Item Response Theory property is discrimination—or the ability of the item to differentiate between people with different strengths of the underlying trait. Ideally, we would like to see items with fairly strong discrimination capturing a range of possible responses dependent on the trait level (Baker, 2017). We used a discrimination value of 1.34 or greater as the cut-point to retain items because that value has been taken as a rule of thumb for identifying items with strong discrimination (Baker, 2017). The range of possible responses can be illustrated in category characteristic curves which indicate the probability that one with a specific trait level will select a response option. Theoretically, people lower in the trait of HCAAP should be more likely to select lower response options than those who are higher in the HCAAP trait.

For the 3-month FFQ, means, standard deviations, and skewness are reported in Table 1. The internal reliability for the full scale was strong, *Cronbach's alpha* = .81, 95% CI [.78, .84]. The Item Response Theory analyses revealed that some items had desirable measurement properties whereas others did not (see Supplementary Materials). Three of the items (dairy, eggs, any meat) had low discrimination (< 0.84), suggesting that those items did not differentiate well those of different levels of the underlying trait (see Supplementary Materials). Even given these problematic items, the test information function (see Supplementary Materials) suggested that the 3-month FFQ provided information across different levels of HCAAP. For subsequent analyses, we removed those three problematic items (dairy, eggs, any meat) which slightly increased the internal reliability, *Cronbach's alpha* = .85, 95% CI [.82, .88].

Some of the items retained in the 3-month measure may appear to overlap in content risking a violation of the local independence assumption in IRT models. To address this, we conducted an analysis of correlation residuals (Yen's Q3) in accordance with Yen

PsychOpen GOLD

(1981) in the Mirt package in R (Chalmers, 2012). For the 3-month FFQ, the mean correlation residual was -.16, *minimum* = -.28 and *maximum* = .07, indicating no substantial violation of local independence (a value of .3 or greater has been used as a rule of thumb for identifying violations of local independence, but no standard rules of thumb have been generally accepted (Christensen et al., 2017). Full Yen's Q3 values are available in the Supplementary Materials.

A similar process was performed for the 24-hour FFQ; *Cronbach's alpha* = .97, 95% CI [.97, .98], means, standard deviations, and skewness are reported in Table 1. The analysis suggested that all the items had strong discrimination (see Supplementary Materials). The category characteristic curves (see Supplementary Materials) suggested that item responses were a function of the strength of the underlying trait. The test information function (see Supplementary Materials) suggested that the 24-hour measure provided the most information for those who consumed many animal products and provided little information about those who consumed few. The results of Yen's Q3 test for local independence did not reveal any substantial violations of the assumption of local independence, *mean* = -0.08, *minimum* = -.26, *maximu*m = .28.

The last step in the planned analyses was to estimate correlations between the two FFQs and the other instruments (see Table 2). The predicted positive relations between the 4Ns and the two FFQs was found. However, there was not a significant relation between the two FFQs and political orientation. Consistent with previous research, those identifying as female, those who were more numerate, and those who knew more about farmed animals reported lower HCAAP scores on both FFQs (A. Feltz & Feltz, 2021; S. Feltz & Feltz, 2019b).

Table 2

Correlations, Means, Standard Deviations, and 95% Confidence Intervals (in Brackets) From Studies 2 (Upper Values) and 3 (Lower Values)

Variable	1	2	3	4	5	6	M	SD
1. 3-month	—						2.98	0.63
							2.61	2.11
2. 24-hour	.66** [.56, .73]	—					2.15	1.23
	.71** [.64, .77]						2.11	1.12
3. KAFS	-.58** [-.65, -.49]	-.64** [-.7, -.56]	—				6.72	1.94
	-.50** [-.59, -.40]	-.61** [-.68, -.52]					6.64	1.82
4. 4Ns	.57** [.49-.65]	.40** [.30, .50]	-.52** [-.60, -.42]	—			4.23	1.27
	.42** [.31, .51]	.42** [.31, .51]	-.40** [-.50, -.30]				4.39	0.84
5. BNT	-.40** [-.49, -.29]	-.44** [-.53, -.33]	.39** [.29, .49]	-.29** [-.4, -.17]	—		2.91	2.03
	-.27** [-.38, -.15]	-.40** [-.40, -.29]	.26** [.14, .37]	-.18** [-.30, -.06]			2.66	1.90
6. Sex	-.24** [-.35, -.12]	-.15* [-.27, -.03]	.26** [.15, .37]	-.35** [-.45, -.24]	-.04 [-.16, .08]	—		
	-.22** [-.33, -.10]	-.24** [-.35, -.12]	.18** [.06, .30]	-.23** [-.35, -.11]	.01 [-.11, .13]			
7. Politics	.08 [-.05, .19]	.06 [-.06, .18]	-.11† [-.23, .01]	.22** [.10, .33]	-.12† [-.23, .01]	.03 [-.10, .15]	3.48	1.79
	.23** [.11, .35]	.23** [.11, .35]	-.22** [-.34, -.10]	.15* [.03, .27]	-.25** [-.37, -.13]	-.04 [-.17, .08]	3.60	1.83

Note. 3-month = 3-month FFQ, 24-hour = 24-hour FFQ, KAFS = Knowledge of Animals as Food scale, 4Ns = 4Ns rationalizations for consuming animals, BNT = Berlin Numeracy Test. Sex was coded 1 = Male, 2 = Female. Politics was measured from 1 (very liberal) to 7 (very conservative).

†*p* < .1. **p < .05. ***p < .01.

Study 3

Study 3 was designed to replicate the results of Study 2 with one modification. The dairy, eggs, and any meat items were removed from the 3-month FFQ because they had low discrimination in Study 2. Since we had good reason to suspect that the items used in Study 3 would constitute the final set of items, we also planned to calculate Item Response Theory model fit statistics for each of the FFQs.

Method

Participants

Two hundred fifty-two participants were recruited from Amazon's Mechanical Turk. Fifty-four percent identified as female (N = 137) and the rest identified as male and the mean age as 35.65, SD = 10.9.

Materials

Participants received the same materials and in the same order as those in Study 2 with one important modification. For the 3-month FFQ, the eggs, dairy, and any meat items were removed. The final set of items and instructions were:

3-Month FFQ — Instructions: How often, in the past 3 months, did you eat the following? (1 = never, 2 = less than 1 time per week, 3 = 1–3 times per week, 4 = 4–6 times per week, 5 = 1 or more times per day).

1. Chicken (fried chicken, in soup, grilled chicken, etc.)
2. Fish and seafood (tuna, shrimp, crab, etc.)
3. Pork (ham, pork chops, ribs, etc.)
4. Beef (steak, meatballs, in tacos, etc.)
5. Bacon
6. Hamburgers

24-hour FFQ — Instructions: In the past day, how many times did you consume the following food? (1 = 0 times, 2 = 1 time, 3 = 2 times, 4 = 3 times, 5 = 4 times, 6 = 5 times or more).

1. Dairy (cheese, milk, yogurt, etc.)
2. Chicken (fried chicken, in soup, grilled chicken, etc.)
3. Turkey (turkey dinner, turkey sandwich, in soup, etc.)
4. Fish and seafood (tune, shrimp, crab, etc.)
5. Pork (ham, pork chops, ribs, etc.)
6. Beef (steak, meatballs, in tacos, etc.)
7. Eggs (omelet, in salad, in baked goods, etc.)

PsychOpen GOLD

8. Bacon
9. Sausages
10. Processed meats (e.g., salami, bologna, etc.)
11. Hamburgers
12. Meat (any type of meat, including beef, pork chicken, turkey, fish, shellfish, or other meats)

Results and Discussion

Analyses proceeded by first examining the item-level properties of the FFQs using a graded response Item Response Theory model (means, standard deviations, and skewness are reported in Table 1). The items in the 3-month FFQ, *Cronbach's alpha* = .81, 95% CI [.77, .84] displayed acceptable item-level properties, like those observed in Study 2 (see Supplementary Materials for analyses, the category characteristic curves, and the test information function). We compared a constrained graded response model that held discrimination among the items fixed (*AIC* = 3649.71, *BIC* = 3737.95) against an unconstrained model (*AIC* = 3595.69, *BIC* = 3701.57). The unconstrained model fit the data better ($p < .01$). A model fit analysis of the three-way margins for the unconstrained model did not reveal any significant model misfit (values < 250.86, *ns*). An exploratory factor analysis using parallel analysis revealed only one factor, supporting the assumption that the items measured only one underlying factor (Factor 1 *eigenvalue* = 2.57, Factor 2 *eigenvalue* = 0.17). There was no substantial violation of local independence: Yen's Q3 *mean* = -.14, *minimum* = -.42, *maximum* = 0.05.

The items in the 24-hour FFQ displayed acceptable item-level properties, similar to those observed in Study 2 (*Cronbach's alpha* = .96, 95% CI = [.95, .97], means, standard deviations, and skewness are reported in Table 1). An unconstrained graded response model (*AIC* = 6437.84, *BIC* = 6691.96) fit the data better than a constrained graded response model (*AIC* = 6495.84, *BIC* = 6710.87), $p < .01$. A model fit analysis of the three-way margins for the unconstrained model did not reveal any significant model misfit (values < 511.5, *ns*). An exploratory factor analysis using parallel analysis revealed only one factor, supporting the assumption that the items measured one underlying factor (Factor 1 *eigenvalue* = 8.13, Factor 2 *eigenvalue* = 0.32). These results largely replicated what was found in Study 2 (see Supplementary Materials for analyses, the category characteristic curves, and the test information function). There was no substantial violation of local independence with a Yen's Q3 *mean* = -.07, *minimum* = -.3, *maximum* = 0.22.

Correlations among the variables gathered are reported in Table 2. The results largely replicated the correlations found in Study 2. The two FFQs were positively correlated. There were moderate to strong negative correlations between knowledge, numeracy, and identifying as female with the FFQs. The FFQs were positively correlated with the 4Ns and with being more politically conservative.

General Discussion

In three studies, we provided some evidence that one can conceive of HCAAP as a trait. Studies 1 and 2 developed items for the 24-hour and 3-month measures of HCAAP. Problematic items were eliminated in Studies 1 and 2, and the remaining items were validated in Study 3. The final items used in Study 3 had acceptable Item Response Theory properties. Studies 2 and 3 also replicated the results of previous research where those who were higher on 4Ns instrument consumed more animal products, and those who were more knowledgeable about animals used as food, were more numerate, or identified as female consumed fewer animal products. In total, the three studies provided evidence that the two FFQs are good candidate measures of trait-like HCAAP behaviors.

Treating HCAAP as a trait can avoid some of the controversies about FFQs in general. One problem with FFQs is how well FFQs capture accurate consumption amounts. Substantial evidence suggests that FFQs are often biased and often do not provide exact estimates of animal or nutrient consumption (Thompson et al., 2015), even if FFQs are often reliable estimators of consumption (Prentice et al., 2011). The same kinds of biases in accuracy of actual HCAAP are likely to apply to our HCAAP FFQs. However, our goal was not to provide accurate measures of true HCAAP amounts. Indeed, our instrument *cannot* provide information about many kinds of food items that are not measured (e.g., liver consumption). Our instruments can, and we think we have evidence they likely do, measure dispositions to consume animals. As such, even though the 24-hour and 3-month measures did not include the same items, they may still be caused by a common disposition to consume animal products. For example, while the 3-month measure did not include prompts about eggs and dairy in the final instrument, those identified as consuming meat and fish are also likely to consume dairy and eggs as well (e.g., in Study 2, those who consumed chicken also tended to consume eggs, $r = .20$, $p < .01$). As such, our measures can give a sense of the magnitude of HCAAP but not an estimate of the quantity of (any specific) animal products consumed. Estimating and measuring reductions in those tendencies can be of value and often are the goal of many researchers even in the absence of accurate, specific information about actual HCAAP.

One may worry that the 24-hour FFQ is not likely to be a useful measure of an HCAAP trait. Reports on 24-hour periods of time are likely to be unstable or reflect other situational factors that are not part of the trait (e.g., maybe one reports food consumption when there was a special celebration the day before). In these ways, the 24-hour measure may not reliably estimate the trait. In a separate, small sample study of college undergraduates ($N = 35$), we gathered data on the 3-month measure and the 24-hour measure 7 days apart (for more information about this study, see Supplementary Materials). The test-retest reliabilities of each instrument were strong: 3-month test, retest $r = .86$, 95% CI [.72, .93]; 24-hour test, retest $r = .83$, 95% CI [.68, .91]. These results suggest that the measures are temporally stable at least in the short term.

PsychOpen GOLD

But still, it is possible that instruments with strong test-retest reliability may not provide evidence for traits or the amount of variation that can be attributed to traits (Geiser & Lockhart, 2012). To help estimate the amount of variance that can be attributed to traits across the two measurement times in that small sample study, we employed Latent State-Trait analysis (Hagemann & Meyerhoff, 2008). In this case and because of the small sample size, we used Bayesian estimates of the amount of variance that can be attributed to underlying traits in both instruments. The 24-hour and 3-month measures both had a large percentage of the variance attributed to traits at both measurement times: 3-month time 1 R^2 = .99, 95% CI [.97, 1.00], 3-month time 2 R^2 = .99, 95% CI [.97, 1.00], 24-hour time 1 R^2 = .91, 95% CI [.72, .99], 24-hour time 2 R^2 = .94, 95% CI [.75, 1.00]. While the point estimates for amount of variance attributed to traits for each measure is high, the small sample size and the 95% CIs warrant some caution, especially for the 24-hour measure. The lower bound 95% CI for the 24-hour measure is much lower than the lower bound for the 3-month measure. This result might indicate that the 24-hour measure is more susceptible to situational or other factors compared to the 3-month measure. As such, there is some reason to think that the 3-month measure is a better measure of the underlying trait of HCAAP. Future research can more fully explore these potential differences between the 3-month and 24-hour measures.

Researchers and activists interested in measuring the degree of HCAAP in their studies may benefit from using the two instruments developed here. But some advice concerning when to use each of the instruments is likely to be beneficial. (Thompson et al., 2015). The test information functions suggest that the two instruments give asymmetric information about the underlying trait of HCAAP. Take the 24-hour FFQ first. In Study 2, only about 16% of the information was captured from thetas (i.e., level of the trait of HCAAP) between -4 and 0 (low consumption) but about 83% of the information was captured in thetas between 0 and 4 (high consumption). In Study 3, there was a similar pattern. From thetas -4 to 0, only about 17% of the information was captured and thetas 0 to 4 captured 81% of the information. However, things were different for the 3-month FFQ. In Study 2, about 44% of the information was captured in thetas -4 to 0, and 41% of the information was captured between thetas 0 and 4. In Study 3, thetas -4 to 0 contained 37% of the information and thetas 0 to 4 contained 58% of the information of the test. So, the 24-hour measure predominately gives information about those who are high on the HCAAP trait and little information about those who are low on the underlying trait. In other words, the 24-hour measure provides information mostly about those who have consumed lots of animal products in a relatively short amount of time. The 3-month FFQ provided more symmetrical information across the spectrum of consumption levels. Because of the limited information provided by the 24-hour measure for low animal product consumers and because of the potential impact of non-trait features on the 24-hour measure, the 3-month instrument is likely to be the better measure of HCAAP as a trait. However, if one's study calls for measurement or

change over a short period of time (e.g., a few days or weeks), then the 24-hour measure may offer some benefit as long as one is aware of the potential measurement issues associated with the 24-hour instrument.

There are several limitations with the current series of studies. First, we want to emphasize that we can only recommend the two FFQs discussed in this paper for research, and not clinical, purposes. More testing on relevant populations would be required to have confidence that the two instruments could be used in clinical settings. Second, in developing the two FFQs, we only gathered data from U.S. IP addresses using one testing platform. The FFQs may behave differently for non-computer literate or non-U.S. residents. A final limitation is that the FFQs proposed here might not offer fine-grained estimates of specific HCAAP behaviors. Rather, our instruments are only designed to give an estimate of the general strength of one's HCAAP tendencies. If one is interested in specific eating behaviors, then lengthier, more detailed FFQs are likely to be better suited to those needs.

Despite these limitations, we think that the short, reliable FFQs reported here offer some advantages to researchers trying to understand and estimate HCAAP tendencies. In general, more work needs to be done to understand the trait of HCAAP, and we hope to have offered one helpful step in that direction.

Supplementary Materials

The supplementary materials provided are the pre-registration of the study, the study items, Yen's Q3 values, FFQ instruments, analyses, power and category characteristic curves, and test information function (for access see Index of Supplementary Materials below):

Index of Supplementary Materials

Feltz, A. (2020). *ACE FFQs* [Pre-registration protocol]. OSF Registries.
    https://doi.org/10.17605/OSF.IO/3DP9M

Feltz, A., Caton, J., Cogley, Z., Engel, M., Feltz, S., Ilea, R., Johnson, L. S. M., & Offer-Westort, T.
    (2023). *Supplementary materials to "Using food frequency questionnaires to measure traits: A case
    study of human consumption of animals and animal products"* [Yen's Q3 values, FFQ
    instruments, analyses, power and category characteristic curves, test information function].
    PsychOpen GOLD. https://doi.org/10.23668/psycharchives.12634

References

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Baker, F. B. (2017). The basics of item response theory using R. Springer.

Buss, D. (1985). The temporal stability of acts, trends, and patterns. In C. Spielberger & J. Butcher (Eds.), *Advances in personality assessment* (pp. 165–196). Erlbaum.

Cade, J., Thompson, R., Burley, V., & Warm, D. (2002). Development, validation and utilisation of food-frequency questionnaires; A review. *Public Health Nutrition, 5*(4), 567–587. https://doi.org/10.1079/PHN2001318

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology, 54*(4), 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q(3): Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. https://doi.org/10.1177/0146621616677520

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1), 25–47. https://doi.org/10.1017/S1930297500001819

Feltz, A., & Feltz, S. (2021). Psychology and vegan studies. In L. Wright (Ed.), *Routledge handbook of vegan studies* (pp. 161–171). Routledge.

Feltz, S., & Feltz, A. (2019a). Consumer accuracy at identifying plant–based and animal–based milk products. *Food Ethics, 4*(1), 85–112. https://doi.org/10.1007/s41055-019-00051-7

Feltz, S., & Feltz, A. (2019b). The Knowledge of Animal as Food Scale. *Human Animal Interactions Bulletin, 7*(2), 19–45. https://doi.org/10.1079/hai.2019.0011

Freedman, L. S., Commins, J. M., Moler, J. E., Arab, L., Baer, D. J., Kipnis, V., Midthune, D., Moshfegh, A. J., Neuhouser, M. L., Prentice, R. L., Schatzkin, A., Spiegelman, D., Subar, A. F., Tinker, L. F., & Willett, W. (2014). Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology, 180*(2), 172–188. https://doi.org/10.1093/aje/kwu116

Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state–trait distinction. *Journal of Personality and Social Psychology, 50*(1), 169–174. https://doi.org/10.1037/0022-3514.50.1.169

Gardner, D., Cummings, L., Dunham, R., & Peirce, J. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement, 58*(6), 898–915. https://doi.org/10.1177/0013164498058006003

Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods, 17*(2), 255–283. https://doi.org/10.1037/a0026977

PsychOpen GOLD

Hagemann, D., & Meyerhoff, D. (2008). A simplified estimation of latent state-trait parameters. *Structural Equation Modeling, 15*(4), 627–650. https://doi.org/10.1080/10705510802339049

Haslam, N., Smillie, L., & Song, J. (2017). *An introduction to personality, individual differences and intelligence* (2nd ed.). SAGE Publications.

Heen, M. S. J., Lieberman, J. D., & Miethe, T. (2014). A comparison of different online sampling approaches for generating national samples. UNLV Center for Crime and Justice Policy. https://doi.org/10.13140/RG.2.2.24283.62243

Kipnis, V., Midthune, D., Buckman, D. W., Dodd, K. W., Guenther, P. M., Krebs-Smith, S. M., Subar, A. F., Tooze, J. A., Carroll, R. J., & Freedman, L. S. (2009). Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics, 65*(4), 1003–1010. https://doi.org/10.1111/j.1541-0420.2009.01223.x

Lance, C. E., Christie, J., & Williamson, G. M. (2021). Do state and trait measures measure states and traits? The case of community-dwelling caregivers of older adults. *Assessment, 28*(3), 829–844. https://doi.org/10.1177/1073191119888582

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment, 25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology, 17*(1), 68–75. https://doi.org/10.1108/02683940210415933

Lusk, J., & Norwood, B. (2016). Some vegetarians spend less money on food, others don't. *Ecological Economics, 130*, 232–242. https://doi.org/10.1016/j.ecolecon.2016.07.005

Myszkowski, N., Storme, M., & Tavani, J. L. (2019). Are reflective models appropriate for very short scales? Proofs of concept of formative models using the Ten-Item Personality Inventory. *Journal of Personality, 87*(2), 363–372. https://doi.org/10.1111/jopy.12395

Piazza, J. M. R., Loughan, S., Luong, M., Kulik, J., Watkins, H., & Seigerman, M. (2015). Rationalizing meat consumption. The 4Ns. *Appetite, 91*, 114–128. https://doi.org/10.1016/j.appet.2015.04.011

Prentice, R. L., Mossavar-Rahmani, Y., Huang, Y., Van Horn, L., Beresford, S. A., Caan, B., Tinker, L., Schoeller, D., Bingham, S., Eaton, C. B., Thomson, C., Johnson, K. C., Ockene, J., Sarto, G., Heiss, G., & Neuhouser, M. L. (2011). Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology, 174*(5), 591–603. https://doi.org/10.1093/aje/kwr140

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25. https://doi.org/10.18637/jss.v017.i05

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*(2), 117–141. https://doi.org/10.1037/bul0000088

Thompson, F. E., Kirkpatrick, S. I., Subar, A. F., Reedy, J., Schap, T. E., Wilson, M. M., & Krebs-Smith, S. M. (2015). The National Cancer Institute's dietary assessment primer: A resource for diet research. *Journal of the Academy of Nutrition and Dietetics, 115*(12), 1986–1995. https://doi.org/10.1016/j.jand.2015.08.016

PsychOpen GOLD

Yen, W. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*(2), 245–262. https://doi.org/10.1177/014662168100500212