Food Quality and Preference 108 (2023) 104880

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

An evaluation and shortening of the Cooking and Food Provisioning Action Scale (CAFPAS) using item response theory

Simon Karlsson [a], Kathryn L. Harris [b], Jeanette Melin [a], Jacob Lahne [c], Julia A. Wolfson [d], Elizabeth S. Collier [b,e,*]

[a] RISE Research Institutes of Sweden, Division of Safety and Transport, Measurement Science and Technology Unit, Gothenburg, Sweden
[b] RISE Research Institutes of Sweden, Division of Bioeconomy and Health, Perception and Design Unit, Stockholm, Sweden
[c] Virginia Polytechnic Institute and State University, Department of Food Science and Technology, Blacksburg, VA, USA
[d] John Hopkins University, Bloomberg School of Public Health, Department of International Health and Department of Health Policy and Management, Baltimore, MD, USA
[e] Linköping University, Division of Society and Health, Department of Health, Medicine and Caring Sciences, Linköping, Sweden

ARTICLE INFO

Keywords:
Cooking ability
Food agency
Item response theory
Categorical measurement
Psychometrics

ABSTRACT

The Cooking and Food Provisioning Action Scale (CAFPAS) is a 28-item validated tool for measuring food agency, a latent construct representing an individual's ability to make and achieve food-preparation and -provisioning goals. Here, key measurement parameters (targeting, threshold ordering, item fit, unidimensionality, differential item functioning, local dependency, and person reliability) of the CAFPAS are evaluated using a specific case of item response theory, Rasch analysis, on data from a development sample (N = 1853; 910 from Sweden; 943 from the US). Winsteps (v.5.1.7) is used for this analysis. The similarity of the Swedish version of the CAFPAS to the original is also assessed. Based on an iterative assessment of the measurement properties with different combinations of items in the development sample, ways to shorten the CAFPAS without jeopardizing construct validity or person reliability are examined. After removing items that do not fit the Rasch model, or that appear redundant in relation to other items, an 11-item version (CAFPAS-short) is suggested and tested using further Rasch analysis on both the development sample and an additional US-based validation sample (N = 1457). Scores of cooking confidence and attitudes are then modelled with measures from the CAFPAS and CAFPAS-short using frequentist and Bayesian analysis. Results suggest that the CAFPAS-short performs similarly to the full-length version, and potential future improvements to the CAFPAS are discussed. This study represents a successful application of item response theory to investigate and shorten a psychometric scale, reducing cognitive load on participants in studies using the CAFPAS whilst minimizing loss of data reliability.

1. Introduction

The sociological concept of food agency describes an individual's ability to plan and execute meal preparation in a variety of contexts, as well as to intentionally make foods as they had envisioned them (Trubek et al., 2017). Despite evidence that lower perceived behavioral control, lack of self-efficacy or confidence regarding food procurement and preparation, and lower food agency can have implications for food choices in terms of both health (Wolfson et al., 2020) and sustainability (Collier et al., 2021, 2022; Graça et al., 2019) there are limited ways to conveniently measure these concepts. This is not surprising, given that quantifying complex latent variables that cannot be measured directly, such as food agency, is not trivial. A review of scales measuring self-reported food agency and cooking literacy among adults indicated that the Cooking and Food Provisioning Action Scale (CAFPAS; Lahne et al., 2017) is among the most comprehensive scales for capturing elements of food agency (Amouzandeh et al., 2019). Speaking to its value as a quantitative tool, CAFPAS scores have been positively associated with healthy dietary intake (Leung et al., 2019; Wolfson et al., 2020), frequency of from-scratch cooking (Wolfson et al., 2020) and number of meals prepared per week (Lahne et al., 2017), as well as inversely related with food neophobia (Niimi et al., 2022). In short, the CAFPAS is a valuable tool for understanding food agency and its relationships with other food-related behaviors.

* Corresponding author at: RISE Research Institutes of Sweden, Division of Bioeconomy and Health, Perception and Design Unit, Stockholm, Sweden.
E-mail address: elizabeth.collier@ri.se (E.S. Collier).

https://doi.org/10.1016/j.foodqual.2023.104880
Received 8 December 2022; Received in revised form 20 April 2023; Accepted 21 April 2023
Available online 26 April 2023

Table 1
Demographics of the development sample (US sample + Swedish sample) and the validation sample.

		US sample* (*N* = 943)	Swedish sample (*N* = 910)	Full development sample (*N* = 1853)	Validation Sample** (N = 1457)
Gender	Female	434 (46.02%)	635 (69.78%)	1069 (57.69%)	738 (50.65%)
	Male	506 (53.66%)	271 (29.78%)	777 (41.93%)	718 (49.28%)
	Other	3 (0.32%)	4 (0.44%)	7 (0.38%)	1 (0.07%)
Age group	18–25	209 (22.16%)	116 (12.75%)	325 (17.54%)	135 (9.27%)
	26–34	346 (36.69%)	234 (25.71%)	580 (31.30%)	261 (17.91%)
	35–54	288 (30.54%)	381 (41.87%)	669 (36.10%)	544 (37.34%)
	55+	100 (10.60%)	172 (18.90%)	272 (14.68%)	514 (35.28%)
	Missing	0 (0.00%)	7 (0.77%)	7 (0.38%)	3 (0.2%)

* data re-used from Lahne et al. (2017).
** data re-used from Wolfson et al. (2020).

The CAFPAS is a 28-item scale consisting of three differentiable subscales, which has been evaluated with classical test theory (CTT) methods: exploratory factor analysis, confirmatory factor analysis, and structural equation modelling (Lahne et al., 2017). The three subscales are: *Self-Efficacy* (one's self-perceived abilities to cook and prepare foods), *Attitude* (one's thoughts on and attitudes about cooking and food provisioning as an activity), and *Structure* (perceived non-food and external barriers to cooking and food provisioning). However, the psychometric properties of these subscales have not yet been validated using item response theory (IRT). Some of the advantages of IRT compared to CTT include the ability to optimize scales for the detection of changes or group differences, increasing scale bandwidth to avoid floor and ceiling effects, to detect and reduce Differential Item Functioning (DIF; which occurs when subgroups systematically respond to items differently after controlling for overall performance), and to examine sensitivity at different levels of the scale (Mallinckrodt & Tekie, 2016). As a result, neglecting IRT methods can lead to a loss of valuable insights regarding the overall functioning of the scale as well as the properties of individual items, which in turn can affect the measure's construct validity and person reliability.

One IRT method for such an investigation of the structure and effectiveness of a scale is Rasch analysis (Rasch, 1960), in which data is evaluated against a measurement model to guide the construction of stable linear measures to assess if requirements for 1) internal validity and 2) an invariant measurement across items and persons are met. Rasch analysis has previously been applied to refine other psychometric scales that had not previously been subject to IRT validations (e.g., Melin et al., 2020; Petrillo et al., 2015; Regnault et al., 2019), as well as to define a hedonic scale for overall liking based on a set of liking ratings from different sensory aspects (Ho, 2019). The Rasch approach can be used to compensate for ordinality in raw data and to enable separate linear measures of person and item attributes on a conjoint linear scale. This makes it possible to, for example, identify which questions in the CAFPAS could contain irregularities (possibly related to the skill of the responders, phrasing, or misinterpretation) as well as those which may be more susceptible to translation issues.

Because poorly translated questionnaires can threaten the validity and usability of a scale (Wild et al., 2005), in addition to following translation and cultural adaptation guidelines, maintaining measurement invariance across languages is critical to ensure reliable and valid comparisons. In practice, this means that the items should work in the same way for the samples from different countries (and for any demographic backgrounds to be compared). If items fail to meet requirements of invariance (i.e., showing DIF), the validity of the comparisons of the person measures may be distorted (Hagquist & Andrich, 2017). The CAFPAS was developed in English and has primarily been used in the US, but translations for other languages and cultures will reasonably follow as it gains traction – as of writing this manuscript the CAFPAS has already been translated to and validated in Czech (Zagata et al., 2022). A Swedish version (analyzed in more depth here) has also been used previously (Niimi et al., 2022).

The present study includes a secondary analysis of the data presented in Lahne et al. (2017) set together with additional CAFPAS data previously collected in Sweden. The primary aim was to evaluate the measurement properties of the CAFPAS using Rasch analysis in order to ascertain its psychometric reliability in detail. In tandem, we evaluated the underlying structure of the Swedish data to assess its similarity to the original English version. Based on an iterative assessment of the measurement properties with different combination of items, we also present shortened versions of the subscales that do not appear to jeopardize validity or reliability. Jaeger & Cardello (2022) state that survey length should be considered, both to minimize boredom and cognitive fatigue, and reduce the risk for missing data (among other suggestions to ensure data quality when collecting survey data, particularly online). Longer surveys can also increase the risk for attrition, which is a concern in itself and may also contribute to sampling bias. At 28 questions, the CAFPAS could be burdensome to include as a module in surveys in some settings, and development of a shorter version (CAFPAS-short) that is less time intensive for respondents without losing data reliability will help improve participant experience and retention in studies employing the CAFPAS as well as broaden its reach.

2. Materials and methods

2.1. Participants and data collection

No new data was collected for this study. Rather, previously collected data from the US and from Sweden as a part of other studies was used. The full development sample consisted of a US sample combined with the Swedish samples: the demographic details are shown in Table 1. The US sample reuses the full dataset collected by Lahne et al. (2017, N = 943). The Swedish data (N = 910) includes one dataset collected in a previously published study (Niimi et al., 2022; N = 389 included in the published manuscript and N = 100 who completed the CAFPAS but were screened out from further participation in that study), and two datasets from unpublished work (N = 96 and N = 325). In all cases, the Swedish CAFPAS data were collected online (during 2020–2021) through advertisement of a survey link on a popular participant recruitment platform. The full development sample was used both for the initial analyses of the full CAFPAS subscales and the derivation of the CAFPAS-short subscales. Data from Wolfson et al. (2020; N = 1457) was used as a validation sample for the CAFPAS-short.

2.2. CAFPAS structure and scoring

The CAFPAS is a 28-item scale consisting of three sub-scales: *Self-Efficacy* (N = 13), *Attitude* (N = 10) and *Structure* (N = 5). Items are answered using 7-point Likert scales ("strongly disagree" =1; "strongly agree" =7). Items that are theorized to be negatively related to the underlying construct are reversed scored (see Table 3). Scores for each subscale are calculated as the standardized sum of responses, and overall raw CAFPAS score is taken as the sum of the three subscale scores (see Lahne et al., 2017).

Table 2
Criteria for the Rasch analyses and their definitions.

Criterion	Description	Criteria/Threshold
Targeting	Rasch analysis allows examination of how well the items' response distributions (thresholds) mirror the abilities of the participants, meaning that there should be a spread in the difficulty of the questions that captures the spread in ability of the respondents.	Item means are conventionally set to a value of 0 logits; thus, *the mean of the corresponding person abilities should be close to 0* in order to claim that they are well represented (Hobart & Cano, 2009).
Threshold ordering	An item threshold represents the point on the latent variable continuum where a respondent has an equal likelihood of choosing either of two adjacent response categories (Andrich, 1978; Wright, & Masters, 1982). It is important for the thresholds of an item to be clearly separated and ordered. Disordered thresholds may indicate that respondents are not able to distinguish between response options and could impact fit statistics and inflate scale reliability (Johansson et al., 2023).	Threshold ordering can be examined by visually inspecting item characteristics curves (Andrich thresholds in Winsteps) where *the thresholds should be separatable and ordered*. If disordered thresholds are identified, however, collapsing adjacent response categories has been highlighted as an effective solution (Van de Winckel et al., 2022).
Item fit	Item fit is quantified using the mean square summary statistics INFIT and OUTFIT, which are both obtained from the squared standardized residual for each item-person interaction (Smith et al., 2008). INFIT emphasizes residuals for items close to the expected value (Stelmack et al., 2004). OUTFIT is sensitive to outliers and accounts for substantial differences between observed and expected values.	If the observed data shall be said to have a *tolerable misfit* (a perfect fit is nearly impossible) with the Rasch measurement model, the mean square summary statistics INFIT and OUTFIT *should both be between 0.7 and 1.3* (Bond & Fox, 2001), where a value of 1.0 would indicate perfect fit.
Unidimensionality	The Rasch model assumes unidimensionality, meaning that the items included in a scale (here, subscales) must measure a single underlying construct (e.g., the *Attitude* subscale measures attitudes towards food preparation).	To support unidimensionality in the subscales, the Eigenvalue *should not be* > 2 in a principal component analysis of the standardized item fit residuals (Boone et al., 2020).
Differential item functioning (DIF)	Fora scale to be considered invariant and generalizable, the items included in the scale should not function differently between groups (e.g., across age groups, genders, languages etc.). In other words, item difficulty should not vary by group characteristics. This is measured as Differential Item Functioning (DIF), which provides an indication of how different groups understand and use the scale. It is not a measure of how raw scores vary across groups. DIF can be considered an interaction effect between an individuals' demographic properties and item (threshold) locations.	A commonly used criterion states that DIF is present *if there is a 0.5 logit or more difference between item calibrations* (e.g., Lai et al., 2003; Pickard et al., 2006).
Local independence	Local independence requires that the correlations between scale items are due to the latent construct that we aim to measure. Thus, if removing the measured construct of interest (e.g., attitude towards cooking), there should be no major associations between any pairs of items included in the scale (Lee, 2004).	For an item pair to be considered locally independent, *there should be no correlations (Yen's Q_3) greater than 0.2* above the average correlations of all scale items between pairs of items (Christensen et al., 2017).
Person reliability	The consistency of the person measures for each sub-scale is evaluated in terms of their reliability, i.e., the proportion of variance that is 'true' variance.	*A person reliability of at least 0.8* indicates good reliability, which corresponds to a person separation index (PSI) of 2.0, and the definition of three statistically different groups of people (Boone et al., 2014; Pendrill & Fisher, 2015).

2.3. Translation procedure for the Swedish CAFPAS

The CAFPAS was first translated by one researcher, who speaks both English (native) and Swedish (high professional proficiency). The translation was checked and refined by one native Swedish speaker. This version was additionally checked by another native Swedish speaker, and edits were made through consultation with other native speakers. The final version (provided in the appendix, Table A1) was translated back to English by a fourth researcher (fluent in Swedish but non-native), after which it was determined that no substantial meaning had been lost during the translation process. The CAFPAS intentionally targets food agency without alluding to culturally specific foods or customs, and so this aspect was not considered strongly during the translation process.

2.4. Data analysis

2.4.1. Exploratory factor analysis (EFA)

To test whether a 3-factor solution for the Swedish data would be comparable to that found for the original US sample (Lahne et al., 2017), the Swedish data were analyzed using EFA targeting three factors with oblique (Promax) rotation and maximum likelihood extraction using *psych::fa* (Revelle, 2016) in R version 4.1.1. (R Core Team, 2019). The purpose of conducting EFA on the Swedish data was two-fold. First, to gain confidence in using it in studies in Sweden where food agency is relevant by ascertaining how similar the Swedish version of the CAFPAS is to the original English version, and second, to affirm the validity of combining the US and Swedish data for subsequent Rasch analysis. EFA with the same parameters was used to assess the underlying structure of the CAFPAS-short for the validation sample.

2.4.2. Rasch analysis

Data were initially handled in Excel and then transferred to Winsteps (v. 5.1.7) for conducting Rasch analyses, and the Partial Credit Model (PCM) was used for all analyses. In the analysis of measurement properties, each subscale was first evaluated against the requirements of fit to the Rasch model. Table 2 provides detailed descriptions of the Rasch criteria that guided the analyses.

Table 3
Results of Rasch analysis on the development sample (N = 1853). Items marked (R) are scored in reverse.

Subscale	Item	Item description	Location (SE)	INFIT MNSQ	OUTFIT MNSQ	DIF
Self-Efficacy	**FSE2**	**I feel limited by my lack of cooking knowledge. (R)**	0.38	1.17	**1.47**	Country: No
			(0.02)			Age: No
						Sex: No
	FSE3	**I can always manage to decide what I would like to eat at any given time.**	0.61	**1.59**	**1.83**	Country: No
			(0.02)			Age: No
						Sex: No
	FSE6	When preparing food, I am confident that I can deal with unexpected results.	0.25	0.88	0.94	Country: No
			(0.02)			Age: No
						Sex: No
	FSE7	**When preparing food, it is easy for me to accomplish my desired results.**	0.18	**0.65**	**0.67**	Country: No
			(0.02)			Age: No
						Sex: No
	FSE8	In preparing food, I can solve most problems with enough effort.	−0.04	0.76	0.78	Country: No
			(0.03)			Age: No
						Sex: No
	ISCO1	**I am comfortable preparing food.**	−0.14	**0.62**	**0.64**	Country: No
			(0.02)			Age: No
						Sex: No
	ISCO5	I know how to use the kitchen equipment I have.	−0.55	0.84	0.80	Country: No
			(0.03)			Age: No
						Sex: No
	ISMP2	**I am involved in daily meal preparation.**	−0.14	1.17	**1.38**	Country: No
			(0.02)			Age: No
						Sex: No
	ISMP3	When I shop for food, I know how I will use the ingredients I am purchasing.	−0.16	1.09	1.26	Country: No
			(0.03)			Age: No
						Sex: No
	ISMP5	I am confident creating meals from the ingredients I have on hand.	−0.11	0.75	0.76	Country: 53
			(0.02)			Age: No
						Sex: No
	ISMP7	**Before I start cooking, I usually have a mental plan of all the steps I will need to complete.**	0.27	**1.57**	**1.81**	**Country: 0.65**
			(0.02)			Age: No
						Sex: No
	ISSH4	When presented with two similar products to purchase, I feel confident choosing between them.	−0.06	1.05	1.09	Country: No
			(0.02)			Age: No
						Sex: No
	ISSH5	I know where to find the ingredients I need to prepare a meal.	−0.48	0.91	0.87	Country: No
			(0.03)			Age: No
						Sex: No
Attitude	**FSE11**	**I find cooking a very fulfilling activity.**	−0.21	**0.62**	**0.61**	Country: No
			(0.02)			Age: No
						Sex: No
	FSE14	For me, cooking is just something to get through as quickly as possible. (R)	0.19	0.75	0.77	Country: No
			(0.02)			Age: No
						Sex: No
	FSE15	Compared to other activities, cooking brings me little enjoyment. (R)	0.08	0.80	0.93	Country: No
			(0.02)			Age: No
						Sex: No
	FSE16	**If I try making a new type of food and it does not come out right, I usually do not try to make it again.** (R)	0.37	**1.55**	**1.69**	Country: No
			(0.02)			Age: No
						Sex: No
	HO4	**I think a lot about what I will cook or eat.**	−0.18	**1.42**	**1.46**	Country: No
			(0.02)			Age: No
						Sex: No

(*continued on next page*)

Table 3 (*continued*)

Subscale	Item	Item description	Location (SE)	INFIT MNSQ	OUTFIT MNSQ	DIF
	HO5	I prefer to spend my time on more important things than food (R)	0.34	0.79	0.83	Country: No
			(0.02)			Age: No
						Sex: No
	ISCO3	**If everything else is equal, I choose to cook rather than have food prepared by someone else.**	0.33	**1.35**	**1.44**	Country: No
			(0.02)			Age: No
						Sex: No
	ISCU4	**I feel like cooking is a waste of effort.** (R)	−0.57	**0.78**	**0.64**	Country: No
			(0.02)			Age: No
						Sex: No
	STR2	I am inspired to cook for other people, like my family or friends.	−0.12	0.97	1.00	Country: No
			(0.02)			Age: No
						Sex: No
	STR3	I feel burdened by having to cook for other people, like my family or friends. (R)	−0.22	1.14	1.18	Country: No
			(0.02)			Age: No
						Sex: No
Structure	**ISMP1**	**I wish that I had more time to plan meals.** (R)	0.99	**1.32**	**1.32**	Country: No
			(0.02)			Age: No
						Sex: No
	STR4	I have a hard time finding enough time to prepare the food I'd like to eat. (R)	0.4	0.83	0.81	Country: No
			(0.02)			Age: No
						Sex: No
	STR8	My family responsibilities prevent me from having time to prepare meals. (R)	−0.73	1.15	1.07	Country: No
			(0.02)			Age: 61
						Sex: No
	STR9	My social responsibilities prevent me from having the time to prepare meals. (R)	−0.64	0.92	0.89	Country: No
			(0.02)			Age: No
						Sex: No
	STR10	My job responsibilities prevent me from having the time to prepare meals. (R)	−0.02	0.87	0.85	Country: No
			(0.02)			Age: No
						Sex: No

Notes. Item measures (location) are presented in logits. Items marked in bold do not satisfy the criteria for item fit specified in Table 2*. INFIT and OUTFIT examines to what degree observed data fit the Rasch model. DIF size is denoted if exceeding the cut-off value of 0.5 logits.*

Person measures are another metric derived from Rasch analysis, which represent respondent ability on the construct being measured expressed on a linear scale. Person measures are obtained from a respondent's raw scores after taking into account the varying difficulties of the items in the test or survey (Boone, 2016). In pursuit of our aim to shorten the subscales, items were excluded if they did not meet the recommended measurement properties (Table 2) or if they were redundant in relation to other items, in an iterative process. To determine if an item was redundant compared to others, we first evaluated the Wright map to assess whether the item and its thresholds captured a unique portion of the latent construct. If an item seemed to be redundant based on the visual inspection of the Wright map, it was removed before re-analyzing the data. We then checked whether the removal of the item resulted in a decrease in PSI. If there was no substantial decrease in PSI, the item was excluded from further analysis.

2.4.3. Relating CAFPAS measures with other factors

Using the validation dataset, raw scores for the three subscales of both the full CAFPAS and CAFPAS-short were calculated, and person measures (see Section 2.4.2) were obtained from these. Scores for cooking confidence, positive attitudes, and negative attitudes to cooking were also calculated (details of measuring these factors are described in Wolfson et al., 2020).

For both the full CAFPAS and CAFPAS-short, Pearson's correlation analysis and multiple linear regression were used to evaluate relationships between person measures for each subscale and scores of cooking confidence and attitudes to cooking. Model performance descriptors (AIC, BIC, and R^2) were obtained for models based on CAFPAS and CAFPAS-short subscales. Person measures were used here because the steps between ordinal rating categories may be unequal if using raw scores, which in turn could skew the results (Boone, 2016). Conditional associations between person measures for each subscale and scores of cooking confidence and attitudes to cooking were also estimated from regression analyses with Bayesian inference (for a comparison of the frequentist and Bayesian approaches, seeBendtsen, 2018) using *rethinking::quap* (McElreath, 2020) in R. Standard normal, priors were used for all coefficients, with $\mu = 0$, $\sigma = 1$, and exponential priors for errors. These priors represent normal distributions centered on 0 (i.e., expected mean influence of subscale person measures of 0 before seeing the data) where 95% of the values land between −2 and 2. The data were centered prior to Bayesian analysis, and means of the posterior distributions provided point estimates for the expected influence of CAFPAS subscale measure on each dependent variable.

3. Results

3.1. EFA of the Swedish CAFPAS

The Swedish sample contained 0.11% missing data. No imputation was made for these, as pattern analysis using *naniar::mcar_test* (Tierney & Cook, 2023) in R indicated that the missing data were Missing Completely at Random, $p = 0.21$. The data were found to be suitable for

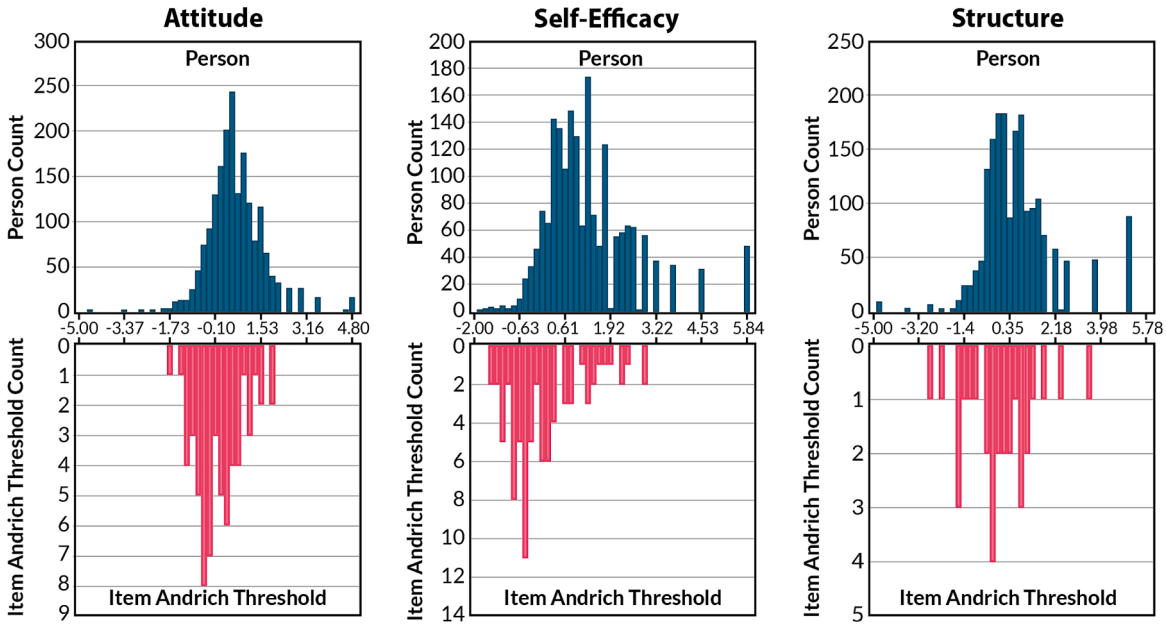

Fig. 1. Person-Item location distributions for each of the three subscales of the CAFPAS from the development sample (N = 1853).

EFA based on the Kaiser-Meyer-Olkin measure of sampling adequacy (0.94) and Bartlett's Test of Sphericity, $\chi^2(378) = 14079.48$, $p < .001$. The scree plot showed a small elbow after the second factor, and a second elbow after the third factor. When three factors were extracted, 49% of the variance was explained, similar to the 47% reported by Lahne et al. (2017). Although a significant goodness-of-fit test indicated imperfect model fit, $\chi 2(297) = 1420.64$, $p < .001$, the RMSEA index (0.065; 90% CI = 0.062 – 0.068) was acceptable and the full Swedish scale showed good internal reliability (Cronbach's α = 0.92). The factor loadings, CAFPAS subscale titles, as well as the items in English are provided in the appendix (Table A1).

The first factor (Cronbach's α = 0.90) corresponded to the *Attitude* subscale from the original English CAFPAS, with all ten items included in the original subscale loading on this factor. Interestingly, the item FSE2 ("I feel limited by my lack of cooking knowledge") also loaded on the first factor, which was unexpected as this item originally loaded on the *Self-Efficacy* subscale (Lahne et al., 2017). The remaining twelve items that loaded on the original *Self-Efficacy* factor did load on the second factor here (Cronbach's α = 0.90 without FSE2; Cronbach's α = 0.89 including FSE2). Finally, the third factor included the five items of the original *Structure* subscale. Correlations between the factors were: *Self-Efficacy* and *Attitude, r* = 0.65, *Self-Efficacy* and *Structure, r* = 0.13, *Attitude* and *Structure, r* = 0.18. In summary, the Swedish version seems to perform very similarly to the original English version, with three underlying factors constructed from near-identical item loading patterns, all of which show good internal reliability. As such, we combined the US and Swedish datasets for subsequent Rasch analysis.

3.2. *Measurement properties of the full CAFPAS in the combined English and Swedish datasets*

Each subscale was evaluated against the requirements of fitting the Rasch model, as described in Table 2 and the results are presented in Table 3. The analyses indicated that all CAFPAS items except HO5 had issues regarding threshold ordering (i.e., the probability curves of the different response options overlapped substantially) which may be a consequence of respondents not being able to reliably differentiate between the seven response options. Additionally, multiple items showed item fit statistics outside the recommended range and some items had problems with DIF. These shortcomings were considered during further analyses (i.e., the generation of the CAFPAS-short subscales). An account of the items affected as well as a summary of the measurement properties (relative to the criteria outlined in Table 2) of each subscale follows, and the Person-Item location distributions are presented in Fig. 1.

Self-Efficacy: A mean person measure of 1.30 (SD = 1.29) indicated a lack of items located towards the higher end of the *Self-Efficacy* subscale (Fig. 1). The subscale showed adequate person reliability of 0.83 (PSI = 2.23). Items in this subscale showed no major problems in terms of unidimensionality or local dependency. However, the analysis of fit statistics showed that the items FSE2, FSE3, FSE7, ISCO1, ISMP2, ISMP5, and ISMP7 were located outside the recommended range, although most only slightly. FSE3 and ISMP7 exhibited the largest deviation from ideal fit. Moreover, ISMP5 and ISMP7 showed a country DIF effect greater than the pre-set 0.5 cut-off value, implying differences in item difficulty between countries.

Attitude: A mean person measure of 0.64 (SD = 0.97) suggested that there was a slight mismatch between person abilities and item difficulties – persons were in general located higher than items (Fig. 1). The overall person reliability of the *Attitude* subscale was adequate, at 0.82 (PSI = 2.14). Items in the *Attitude* subscale had no major issues in regard to unidimensionality, DIF, or local dependency. Regarding item fit, the items FSE11, FSE16, and HO4 were located outside the recommended range, although most only slightly.

Structure: A mean person measure of 0.73 logits (SD = 1.47) indicated that participants, in general, were located higher than items on the shared logit scale (Fig. 1). The *Structure* subscale showed an acceptable, though slightly low, person reliability of 0.79 (PSI = 1.92). This subscale showed no major problems in terms of unidimensionality or local dependency, but the item ISMP1 showed a slight underfit. DIF-analysis showed that the item STR8 ("My family responsibilities prevent me from having time to prepare meals") had slight-to-moderate DIF effects for age and gender (0.61 logits), signifying that the item difficulty varied depending on the age and gender of the participants. The largest contrast for age was found between the age groups 18–25 and 35–54, suggesting that participants in the typical age of parenthood were more likely to endorse the item given a certain position on the overall scale.

Table 4
Exploratory factor analysis using the validation sample (N = 1457) on the suggested items for the CAFPAS-short following iterative analyses of the development sample.

Subscale	Item	Item description	Factor 1 (Self-Efficacy)	Factor 2 (Attitude)	Factor 3 (Structure)
Self-Efficacy	FSE6	When preparing food. I am confident that I can deal with unexpected results.	**0.77**	−0.02	−0.01
	FSE7	When preparing food. It is easy for me to accomplish my desired results	**0.83**	−0.02	0.01
	FSE8	In preparing food. I can solve most problems with enough effort.	**0.80**	−0.06	−0.01
	ISCO1	I am comfortable preparing food.	**0.73**	0.14	0.00
Attitude					
	FSE14	For me, cooking is just something to get through as quickly as possible. (R)	0.02	**0.83**	−0.03
	FSE15	Compared to other activities. Cooking brings me little enjoyment. (R)	−0.03	**0.68**	−0.06
	HO5	I prefer to spend my time on more important things than food (R)	−0.03	**0.75**	0.05
	ISCU4	I feel like cooking is a waste of effort. (R)	0.08	**0.62**	0.13
Structure	STR4	I have a hard time finding enough time to prepare the food I'd like to eat. (R)	0.08	0.09	**0.69**
	STR9	My social responsibilities prevent me from having the time to prepare meals. (R)	−0.07	0.03	**0.79**
	STR10	My job responsibilities prevent me from having the time to prepare meals. (R)	−0.02	−0.08	**0.90**

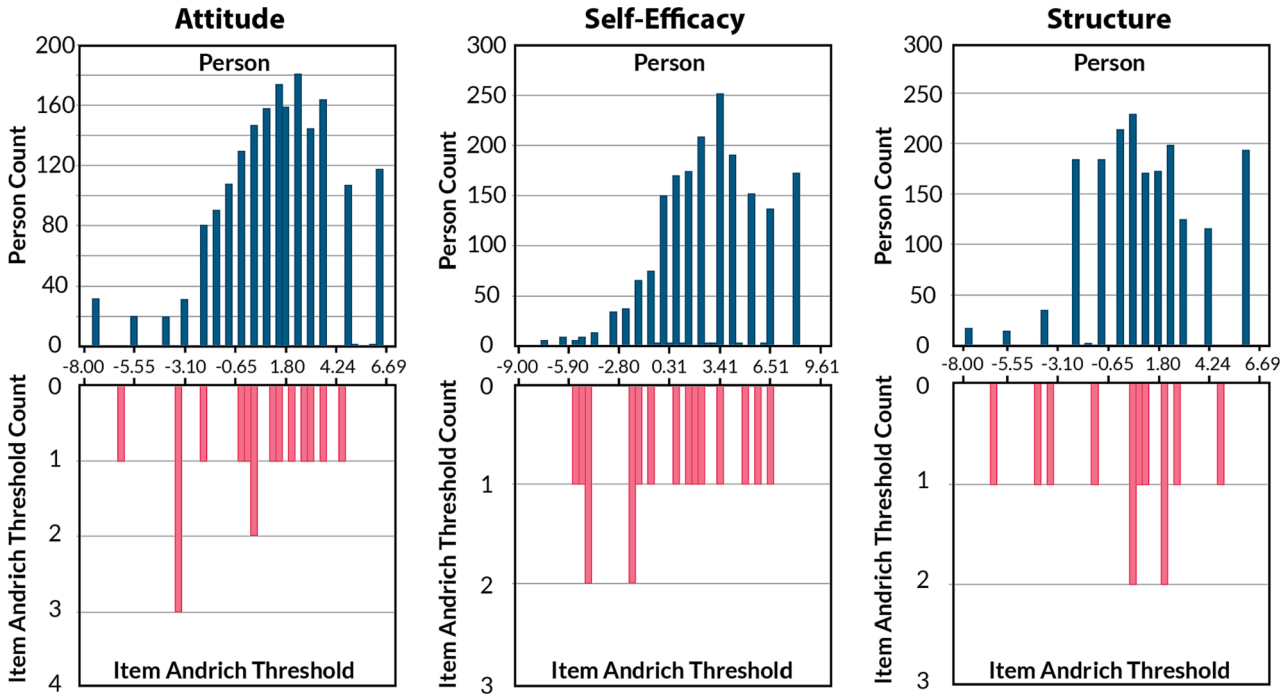

Fig. 2. Person-Item location distributions for each of the three subscales of the CAFPAS-short for the development sample (N = 1853).

3.3. Development of the CAFPAS-short

Having assessed the internal psychometric properties of the full CAFPAS scale, we went on to test the feasibility of shortening the scale using an iterative assessment of the measurement properties with different combination of items within each subscale. The resulting CAFPAS-short (see Table 4) consisted of 11 items (*Self-Efficacy* N = 4; *Attitude* N = 4; *Structure* N = 3). The Person-Item distributions are provided in Fig. 2. Because the items included in the CAFPAS-short subscales based on the development sample still displayed some disordered thresholds, the original 7 response options were merged into 5 ordinal categories throughout the dataset according to the following rule: 1 = 1, 2 = 2, 3 = 2, 4 = 3, 5 = 3, 6 = 4, 7 = 5.

The mean person measures for the subscales of the CAFPAS-short (based on the development sample) were 1.18 logits (SD = 2.60) for *Structure*, 1.35 logits (SD = 2.66) for *Attitude*, and 2.72 logits (SD = 3.03) for *Self-Efficacy*. The reliability statistics of the shortened subscales were nearly the same as the original scales: 0.77 (PSI = 1.84) for *Structure*, 0.84 (PSI = 2.27) for *Attitude*, and 0.84 (PSI = 2.26) for *Self-Efficacy*. In accordance with the targeting analyses of the full CAFPAS subscales, there were no items with difficulty levels appropriate for capturing participants located towards the higher end of the subscales (Fig. 2). Additionally, three items (FSE8, FSE14, FSE15) showed a country DIF larger than the specified cut-off value, which could be a consequence of cultural differences or translation effects – these issues are addressed in the discussion section. The item ISCO1 also showed a DIF effect for age, but because the effect was just above the cut-off value, the item was retained for the validation analysis in order to examine whether the DIF remained. None of the shortened subscales showed problems when it came to threshold ordering, item fit, unidimensionality, or local dependency. As such, we went on to examine whether the results could be validated in an independent sample and compared scores from the full CAFPAS and CAFPAS-short in terms of their relationships with cooking confidence and attitudes to cooking.

Table 5
Results of Rasch analysis of the CAFPAS-short for the development and validation samples.

Subscale	Item	Item description	Location (SE)		INFIT MNSQ		OUTFIT MNSQ		DIF	
			Development	Validation	Development	Validation	Development	Validation	Development	Validation
Self-Efficacy	FSE6	When preparing food. I am confident that I can deal with unexpected results.	0.60	0.61	1.08	1.08	1.08	1.07	Country: No	Age: No
			(0.05)	(0.05)					Age: No Sex: No	Sex: No
	FSE7	When preparing food. it is easy for me to accomplish my desired results	0.38	0.08	0.99	0.92	0.99	0.90	Country: No	Age: No
			(0.05)	(0.05)					Age: No Sex: No	Sex: No
	FSE8	In preparing food. I can solve most problems with enough effort.	−0.28	0.10	1.00	1.05	0.99	1.05	**Country: 0.61**	Age: No
			(0.05)	(0.05)					Age: No Sex: No	Sex: No
	ISCO1	I am comfortable preparing food.	−0.69	−0.79	0.87	0.89	0.83	0.90	Country: No	Age: No
			(0.05)	(0.05)					**Age: 0.56** Sex: No	Sex: No
Attitude	FSE14	For me. cooking is just something to get through as quickly as possible.	0.60	0.24	1.01	0.84	1.00	0.83	**Country: 0.60**	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No
	FSE15	Compared to other activities. cooking brings me little enjoyment.	0.26	0.60	1.03	1.29	0.98	1.22	**Country: 0.78**	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No
	HO5	I prefer to spend my time on more important things than food	1.00	0.41	1.01	1.01	0.99	0.98	Country: No	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No
	ISCU4	I feel like cooking is a waste of effort.	−1.85	−1.25	0.94	0.83	0.900	0.79	Country: No	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No
Structure	STR4	I have a hard time finding enough time to prepare the food I'd like to eat.	1.10	0.63	1.09	1.12	1.05	1.10	Country: No	Age: No
			(0.04)	(0.05)					Age: No Sex: No	Sex: No
	STR9	My social responsibilities prevent me from having the time to prepare meals.	−1.21	−0.57	1.04	0.99	1.02	0.99	Country: No	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No
	STR10	My job responsibilities prevent me from having the time to prepare meals.	0.11	−0.06	0.83	0.86	0.81	0.84	Country: No	Age: No
			(0.04)	(0.04)					Age: No Sex: No	Sex: No

Notes. Item measures (location) are presented in logits. INFIT and OUTFIT examines to what degree observed data fit the Rasch model. DIF size is denoted if exceeding the cut-off value of 0.5 logits. All thresholds were organized after reducing the original seven response options into five.

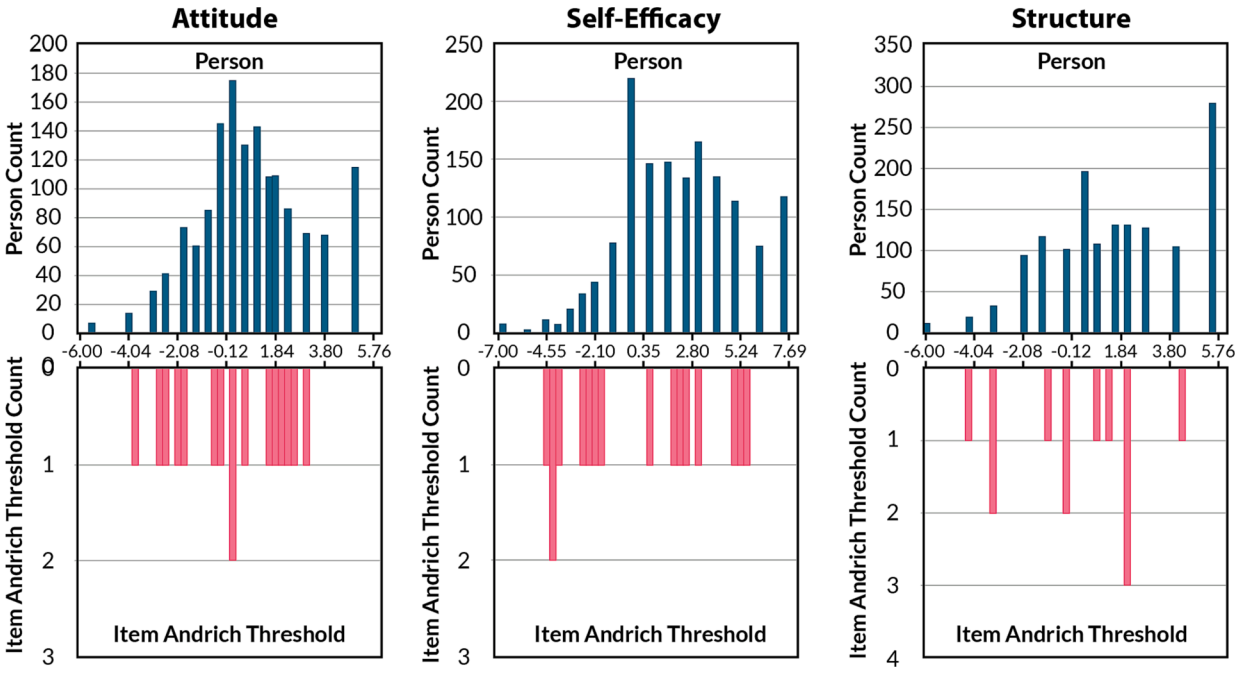

Fig. 3. Person-Item location distributions for each of the three subscales of the CAFPAS-short for the validation sample (N = 1457).

Table 6
Pearson correlations (r) and standardized regression coefficients (β), shown as r(β), as well as model performance descriptors (AIC, BIC, and R^2) values from multiple linear regression with subscale person measures as covariates for CAFPAS and CAFPAS-short subscales and cooking confidence and attitudes.

Scale	Subscale	Cooking confidence	Positive cooking attitudes	Negative cooking attitudes
CAFPAS	**Self-Efficacy**	0.67*** (0.59***)	0.61*** (0.40***)	-0.51*** (-0.07**)
	Attitude	0.54*** (0.18***)	0.63*** (0.46***)	-0.66*** (-0.54***)
	Structure	0.11*** (-0.15***)	-0.042 (-0.33***)	-0.43*** (-0.22***)
	AIC	3928.1	3324.0	4030.5
	BIC	3954.5	3350.5	4056.9
	R^2	0.47	0.55	0.49
CAFPAS-short	**Self-Efficacy**	0.73*** (0.70***)	0.69*** (0.63***)	-0.48*** (-0.20***)
	Attitude	0.39*** (0.11***)	0.44*** (0.30***)	-0.70*** (-0.52***)
	Structure	0.17*** (-0.09***)	-0.06* (-0.28***)	-0.50*** (-0.16***)
	AIC	3735.9	3312.7	3868.5
	BIC	3762.3	3393.2	3894.9
	R^2	0.54	0.55	0.54

*** $p < .001$, ** $p < .01$, * $p < .05$.

3.4. Evaluating the CAFPAS-short

3.4.1. EFA on the validation sample

Response options for the validation sample were also merged as described in Section 3.3. EFA conducted on the items included in the CAFPAS-short for the validation sample showed that all items loaded on the expected subscale (Table 4). Thus, a three-factor solution showing the expected construction, *Self-Efficacy* (Cronbach's α = 0.87), *Attitude* (Cronbach's α = 0.83), and *Structure* (Cronbach's α = 0.84) and explaining 61% of the variance in the data was obtained. *Self-Efficacy* correlated positively with *Attitude* ($r = 0.55$) and *Structure* ($r = 0.31$), as did *Attitude* and *Structure* ($r = 0.63$).

3.4.2. Rasch analysis of the CAFPAS-short using the validation sample

A separate Rasch analysis was conducted on the validation data to examine whether the subscales worked similarly in an independent sample. For the validation analysis to be considered successful, the data should again meet the requirements of fitting the Rasch model (see Table 2). Additionally, the item locations in the validation sample should not deviate substantially (more than 0.5 logits) from the corresponding item location in the development sample. The results of Rasch analysis of the CAFPAS-short for both the development and validation samples are shown in Table 5.

In all, results similar to those from the development sample were obtained using the validation sample in terms of item fit statistics and DIF. The *Self-Efficacy* scale did not show a substantial DIF effect for age in the validation sample and did not show any deviances in item location above the cut-off value. For the *Attitude* scale, a slightly different item hierarchy was obtained in the validation sample compared to the development sample and the items HO5 and ISCU4 deviated more than 0.5 logits compared to the development sample. For instance, the item HO5 were ranked the most difficult item (i.e., hardest to agree on) in the development sample, whereas it was the second most difficult item in the validation sample. Finally, the *Structure* scale displayed a similar item hierarchy as in the development sample, although the item location for STR9 deviated 0.64 logits between the two samples. Person-Item location distributions are presented in Fig. 3.

3.4.3. Relationships with cooking confidence and attitudes

Pearson correlation coefficients as well as standardized regression coefficients and descriptors of model performance from multiple linear regression on cooking confidence and attitudes (positive and negative)

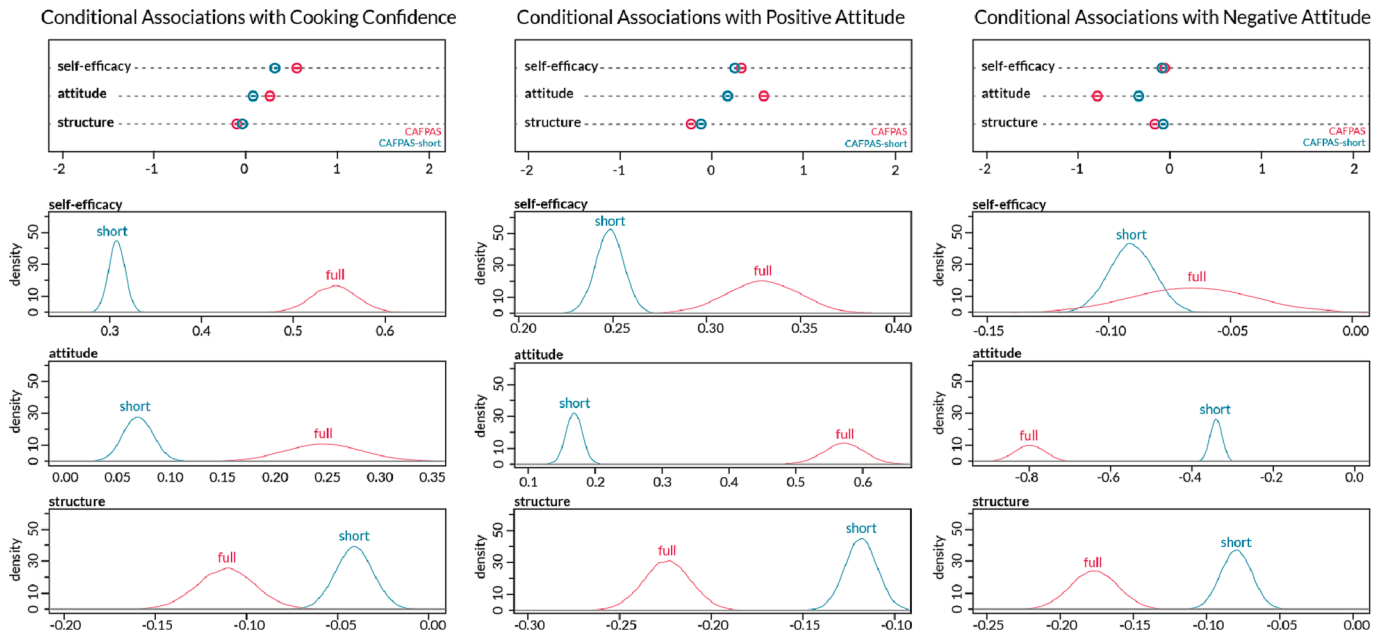

Fig. 4. Shows the conditional associations for the person measures of Self-Efficacy, Attitude, and Structure with positive cooking attitudes, negative cooking attitudes, cooking confidence, and cooking behavior. Top row: The points show the means of the posterior distributions and indicate the most probable change in the other variables given a one-point increase in Self-Efficacy, Attitude, or Structure, as calculated for the full CAFPAS (pink) and CAFPAS-short (green) using the validation sample. Error bars show the 89% compatibility intervals. Remaining rows: Posterior distributions for each case. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using subscale person measures from the CAFPAS and CAFPAS-short as covariates are shown in Table 6. In all cases, significant models with similar R^2 values are obtained for both CAFPAS and CAFPAS-short, suggesting comparable relationships and model performance.

The posterior distributions and their means, based on modelling the person measures with cooking confidence and attitudes using Bayesian inference (see Section 2.4.3), are shown in Fig. 4. Relationships with cooking attitudes and confidence are directionally the same with both the CAFPAS and CAFPAS-short subscales, although, in agreement with the analyses shown in Table 6, the strength of the relationships differs somewhat. Inspection of the posterior distributions shows that the estimated associations with subscale person measures from the full CAFPAS on cooking attitudes and confidence are often stronger with a wider range, whilst for the CAFPAS-short, expected associations are typically weaker and within a narrower range. Taken together, these analyses suggest that the subscales of the CAFPAS and CAFPAS-short values are associated with other cooking measures in comparable ways.

4. Discussion

Quantifying complex latent variables that cannot be directly measured is challenging. The CAFPAS (Lahne et al., 2017), comprised of three subscales (*Self-Efficacy, Attitude,* and *Structure*), is a one such tool for measuring food agency. Here, we investigated the psychometric properties of these subscales using Rasch analysis, to assess the performance of both individual items and the subscales overall. Because we used data from the US and from Sweden, we additionally investigated the underlying structure of the Swedish version of the CAFPAS using EFA to compare it to the original English version. After evaluating the full CAFPAS, we iteratively tested different combinations of items between reassessments using Rasch analysis to generate a shortened version of the scale (CAFPAS-short), which was then assessed with Rasch analysis and EFA using a separate set of previously collected data (Wolfson et al., 2020). Using this same validation data, we evaluated how CAFPAS measures from the full scale and the CAFPAS-short related to self-reported cooking confidence and attitudes (from a separate survey in the same previous study; Wolfson et al., 2020) using frequentist and Bayesian analysis.

We found that 1) the full CAFPAS is shown to be fit for purpose according to Rasch analysis, 2) the Swedish version of the CAFPAS is comparable in structure to the original English version and captures the same constructs, and 3) the CAFPAS-short is psychometrically sound and produces measures which relate to other variables of interest in a similar way to those derived from the full scale. The remainder of this section addresses areas where there is room for improvement in the full CAFPAS and CAFPAS-short, and routes to addressing this.

Overall, the original CAFPAS performed adequately according to the criteria defined for the Rasch analysis (Table 2). Nonetheless, the analysis revealed that there exist aspects which could be improved. First, an analysis of thresholds indicated that most items had disordered thresholds when using the seven response options used in the original CAFPAS, which could be a consequence of the respondents not being able to reliably discriminate all response options from each other. If not resolved, disordered response categories can impact item fit statistics and overestimate person reliability. Second, multiple items showed an item fit outside the recommended range, raising questions concerning their usefulness in the scale. Third, items that could be described as "difficult" for participants to endorse were lacking (person means were substantially greater than 0, see Fig. 1). This, in turn, leads to ceiling effects and poorer discrimination between participants with higher cooking-related abilities. Finally, the interpretation of some individual items (ISM5, ISMP7, STR8) seemed influenced by demographic characteristics. This implies that the relative difficulty level of these items may vary as a consequence of group characteristics, which could pose a threat to the invariance and generalizability of the scale (Table 2).

Rasch analysis conducted on the CAFPAS-short indicated that it performs well and similarly to the full CAFPAS, making it a good option for studies where a shorter scale may be more desirable than using the full 28-item scale. Importantly, by systematically evaluating the measurement properties with different combinations of items in a Rasch analysis, some of the problems identified for the full CAFPAS described above were resolved. For example, issues with disordered thresholds disappeared after merging the original seven response options into five categories when analyzing the CAFPAS-short, and model fit (according

to the Rasch guidelines, Table 2) was improved by excluding items that did not fit the Rasch model. We suggest that future users of the CAFPAS use a 5-point instead of 7-point response scale based on this. However, addressing ceiling effects by authoring new questions, for example, was outside of the scope of this work as no new data was collected.

Some items retained in the CAFPAS-short still displayed country differences according to DIF analysis (development sample only), suggesting some differences in interpretation between the US and Swedish samples. This DIF effect could also explain why some deviances were found in both item hierarchy and item location were present in the development sample but not the validation sample (which included only US participants). Although non-ideal, it seems reasonable that some translation-related shifts in meaning could occur when adapting categorical measurements scale to different languages. Importantly, such subtle differences may not be obvious through employing EFA or similar analyses alone, indeed here the EFA results indicated good agreement in underlying structure between the original English version and the Swedish version. This further highlights the added value of employing IRT methods when working with questionnaire and categorical data.

In the Swedish data, FSE2 *("I feel limited by my lack of cooking knowledge")* loaded on the *Attitude* subscale when it was expected to load on the *Self-Efficacy* subscale. Comparisons with the Czech translation of the CAFPAS (Zagata et al., 2022) suggest that this item may be somewhat tricky to translate. Although in Zagata et al. (2022), FSE2 did load on *Self-Efficacy* (factor loading = 0.466), its loading was nearly identical for the *Attitude* subscale (factor loading = 0.464). It is possible that subtle language differences affect the interpretation of this item, shifting it from leaning towards a theme of individual skill and perceived self-ability in English to instead leaning towards a theme of attitudes towards the consequences of lower cooking knowledge in other languages. However, this remains speculative, and it should be noted that we did not observe any country-related differences in this item according to Rasch analysis.

It is encouraging that similar relationships for person measures from the full CAFPAS and CAFPAS-short with cooking confidence and attitudes were observed. These relationships were not identical, however. It is possible that individual items in the full versions of the subscales with stronger relationships with these outcomes may have been excluded in the CAFPAS-short. For example, for negative cooking attitudes, because the items in the short *Attitude* scale only includes reverse score items this could mean that items strongly driving the negative association between negative attitudes and full CAFPAS measures may have been removed, giving a weaker pattern of association. Estimates of subscale influence were typically weaker for the CAFPAS-short compared to the full CAFPAS, but the range of estimated coefficients was often narrower (more precise) for the CAFPAS-short. Without further investigation it remains difficult to ascertain which gives the more accurate picture – e.g., does the CAFPAS-short downplay these associations, or does the full CAFPAS exaggerate them?

Typically, researchers have focused on the relationships between overall CAFPAS scores and other factors (e.g., Lahne et al., 2017; Niimi et al., 2022; Wolfson et al., 2020), as opposed to assessing the relative influence of each subscale. Here, we have instead focused on subscale performance because of the unidimensionality criterion for Rasch analysis, which a composite CAFPAS score calculated from distinct subscales would, by definition, defy. Significant regression models with comparable performance in terms of AIC, BIC, and R^2 were obtained when using the subscales from the CAFPAS and CAFPAS-short as covariates. This shows that multiple regression using subscale measures as covariates is a viable complementary approach to calculating a single CAFPAS score that can facilitate a more detailed understanding of which aspects of food agency are most relevant for assessing different food and cooking behaviors. For example, here we observe that cooking confidence seems more strongly associated with *Self-Efficacy* than *Attitude* or *Structure*. Though out of scope for the present work, assessing the relative influence of each subscale for a suite of food behaviors may highlight where different facets of food agency are of most relevance, and support more tailored guidance of how to support behavioral change. Given that models using the subscales from the long and short versions were highly similar, the CAFPAS-short could be useful in such a pursuit.

4.1. Limitations & future research directions

There are a number of limitations that should be noted. Because both the full CAFPAS and the CAFPAS-short were demonstrated by Rasch analysis to be weaker in their description of people who score highly, a logical next step is to revisit the original design of the CAFPAS. This might include analysis of responses to questions in the original development survey to identify more 'difficult' questions to previously excluded that could be beneficial to reintroduce into the scale. Alternatively, authoring new items on which even high food agency respondents might rate themselves somewhat lower could be relevant. Indeed, the generation of the CAFPAS-short offers an opportunity to revisit originally rejected items, which could be added to the CAFPAS-short without increasing the item count beyond a reasonable length.

The Swedish version of the CAFPAS did perform similarly compared to prior studies in the US, and we conclude it can be used in Swedish-speaking situations. Nonetheless, the translation protocol employed here did not include translation or linguistic experts, instead relying on native and fluent Swedish speakers using a back-translation approach, which, while not without precedent is also not the most highly recommended method. More work is needed to explore food agency in different contexts more generally, particularly outside of the US and with lower-income groups. Though the validation work undertaken in the present study is promising, further revision of the CAFPAS, particularly related to different structural factors in different cultural contexts may be warranted.

The CAFPAS-short was not tested in isolation, instead we used a subset of previously published data to validate it. It has been suggested that participants' responses to a questionnaire item can be influenced by their responses to previous items (Shimada & Katahira, 2022) such that a person's response to a given question tends to be biased towards their response on the preceding question. Here, this could mean that the measures calculated for the CAFPAS-short may have been somewhat different than if only items from the CAFPAS-short had been posed. Future work using the CAFPAS-short as an instrument in its own right will facilitate better understanding of how its subscales may relate to other cooking-related measures.

5. Conclusions

Here we present a secondary analysis of the CAFPAS data reported by Lahne et al. (2017) using item response theory in order to investigate the characteristics of the CAFPAS in depth as well as attempt to generate a shortened version of the scale (CAFPAS-short). We found that the full CAFPAS overall performs adequately according to the Rasch approach, although there could be room for improvement as the subscales seem to experience disordered thresholds, ceiling effects, item fit issues, and item difficulty seemed affected by demographic characteristics in some cases. The CAFPAS-short, however, did not have the same issues with item fit or disordered thresholds, and represents a successful application of IRT to reduce and refine a psychometric scale. The Swedish version of the CAFPAS captured similar constructs, and was comparable in structure to, the original English version, making it suitable for use when Swedish speakers may prefer to use Swedish in studies utilizing the CAFPAS. Additional Rasch analysis on data from Wolfson et al. (2020) to assess and validate the measurement properties of the CAFPAS-short indicated good agreement with the full version. The CAFPAS-short also produced subscale measures which related to other cooking-related variables in a similar way to those produced from the full scale. The CAFPAS-short shows promise as a tool for measuring food agency without losing data reliability and may be particularly useful when the CAFPAS is used as a component of a battery of surveys.

Table A1
Results of exploratory factor analysis on the Swedish (N = 910) data. Original items in English provided in parentheses.

Subscale	Item	Code	Factor 1 (Attitude)	Factor 2 (Self-Efficacy)	Factor 3 (Structure)
Self-Efficacy	Jag känner mig begränsad av mina matlagningskunskaper (I feel limited by my lack of cooking knowledge)	FSE2	**0.43**	0.23	0.09
	Jag kan alltid bestämma mig för vad jag vill äta, när som helst (I can always manage to decide what I would like to eat at any given time)	FSE3	0.18	**0.26**	0.14
	När jag lagar mat, känner jag mig säker på att jag kan hantera oväntade resultat (When preparing food, I am confident that I can deal with unexpected results)	FSE6	0.12	**0.58**	0.08
	När jag lagar mat, är det lätt för mig att uppnå resultatet som jag vill ha (When preparing food, it is easy for me to accomplish my desired results)	FSE7	0.16	**0.73**	−0.01
	När jag lagar mat, kan jag lösa de flesta problem om jag försöker tillräckligt mycket (In preparing food, I can solve most problems with enough effort)	FSE8	0.07	**0.72**	−0.02
	Jag känner mig bekväm med matlagning (I am comfortable preparing food)	ISCO1	0.32	**0.63**	−0.05
	Jag vet hur man ska använda de köksredskap jag har (I know how to use the kitchen equipment I have)	ISCO5	−0.10	**0.76**	0.01
	Jag är involverad i den dagligen matlagningen (I am involved in daily meal preparation)	ISMP2	0.12	**0.53**	−0.03
	När jag handlar mat, vet jag hur jag kommer använda ingredienserna som jag köper (When I shop for food, I know how I will use the ingredients I am purchasing)	ISMP3	−0.12	**0.7**	0.02
	Jag känner mig trygg i att laga måltider utifrån de ingredienser som jag har till hands (I am confident creating meals from the ingredients I have on hand)	ISMP5	0.00	**0.85**	−0.02
	Jag brukar ha en mental plan för alla steg som behövs innan jag börja laga mat (Before I start cooking, I usually have a mental plan of all the steps I will need to complete)	ISMP7	0.04	**0.36**	−0.08
	Vid inköp känner jag mig säker i att kunna välja mellan två liknande produkter. (When presented with two similar products to purchase, I feel confident choosing between them)	ISSH4	−0.08	**0.69**	0.05
	Jag vet var jag kan hitta ingredienserna jag behöver för att laga en måltid (I know where to find the ingredients I need to prepare a meal)	ISSH5	−0.20	**0.74**	0.06
Attitude	Jag tycker att det är väldigt tillfredsställande att laga mat (I find cooking a very fulfilling activity)	FSE11	**0.75**	0.19	−0.12
	Enligt mig, är matlagning bara något att få avklarat så fort som möjligt (For me, cooking is just something to get through as quickly as possible)	FSE14	**0.89**	−0.16	0.06
	Jämfört med andra aktiviteter, gillar jag inte matlagning särskilt mycket (Compared to other activities, cooking brings me little enjoyment)	FSE15	**0.99**	−0.15	−0.01
	Om jag försöker laga något nytt som inte blir lyckat, brukar jag oftast inte försöka laga det igen (If I try making a new type of food and it does not come out right, I usually do not try to make it again)	FSE16	**0.47**	−0.05	0.11
	Jag tänker ofta på vad jag kommer tillaga eller äta (I think a lot about what I will cook or eat)	HO4	**0.42**	0.10	−0.12
	Jag föredrar att spendera tid på viktigare saker än matlagning (I prefer to spend my time on more important things than food)	HO5	**0.98**	−0.23	0.06
	Om allting annat är jämt, skulle jag välja att laga mat själv i stället för att låta någon annan laga den åt mig (If everything else is equal, I choose to cook rather than have food prepared by someone else)	ISCO3	**0.44**	0.09	−0.11
	Jag tycker att matlagning är ett slöseri med energi (I feel like cooking is a waste of effort)	ISCU4	**0.8**	−0.02	0.03
	Jag känner mig inspirerad att laga mat till andra, till exempel min familj eller vänner (I am inspired to cook for other people, like my family or friends)	SRT2	**0.66**	0.13	−0.13
	Jag känner mig belastad av att behöva laga mat till andra, till exempel min familj eller vänner (I feel burdened by having to cook for other people, like my family or friends)	STR3	**0.56**	0.07	0.11
Structure	Jag önskar att jag hade mer tid att planera måltider (I wish that I had more time to plan meals)	ISMP1	−0.19	0.03	**0.63**
	Det är svårt för mig att hitta tillräckligt med tid för att laga maten som jag skulle vilja äta (I have a hard time finding enough time to prepare the food I'd like to eat)	STR4	0.08	−0.04	**0.81**
	Mitt familjeansvar förhindrar mig från att hitta tid till matlagning (My family responsibilities prevent me from having time to prepare meals)	STR8	0.12	0.00	**0.61**
	Mitt sociala ansvar förhindrar mig från att hitta tid till matlagning (My social responsibilities prevent me from having the time to prepare meals)	STR9	0.03	0.06	**0.68**
	Mitt jobbansvar hindrar mig från att hitta tid till matlagning (My job responsibilities prevent me from having the time to prepare meals)	STR10	0.01	0.00	**0.71**

Funding

SK, KLH, JM, and ESC were supported by the RISE competence platform *Centre for Categorical Based Measurements*. KLH and ESC were supported by FORMAS - Research Council for Environment, Agricultural Sciences and Spatial Planning, grant number 2018–01867. JAW was supported by the National Institutes of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH), grant number K01DK119166.

CRediT authorship contribution statement

Simon Karlsson: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Kathryn L. Harris:** Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Jeanette Melin:** Conceptualization, Methodology, Validation, Writing – review & editing, Project administration. **Jacob Lahne:** Conceptualization, Writing – review & editing, Validation, Resources. **Julia A. Wolfson:** Conceptualization, Writing – review & editing, Validation, Resources. **Elizabeth S. Collier:** Conceptualization, Methodology, Formal analysis,

Writing – original draft, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors acknowledge Anne Normann for work with the translation of the CAFPAS to Swedish as well as involvement in collecting data in Sweden. Jun Niimi, Victoria Sörensen, Lisa-Maria Oberrauter, and Penny Bergman are also acknowledged for data collection in the Swedish sample.

Appendix A

References

Amouzandeh, C., Fingland, D., & Vidgen, H. A. (2019). A Scoping Review of the Validity, Reliability and Conceptual Alignment of Food Literacy Measures for Adults. *Nutrients, 11*(4), E801. https://doi.org/10.3390/nu11040801

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573. https://doi.org/10.1007/BF02293814

Bendtsen, M. (2018). A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. *Journal of Medical Internet Research, 20*(10), e10873.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences* (pp. xxiv, 255). Lawrence Erlbaum Associates Publishers.

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education, 15*(4). https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., & Staver, J. R. (2020). Principal Component Analysis of Residuals (PCAR). In W. J. Boone, & J. R. Staver (Eds.), *Advances in Rasch Analyses in the Human Sciences* (pp. 13–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-43420-5_2.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch Analysis in the Human Sciences. *Springer, Netherlands.*. https://doi.org/10.1007/978-94-007-6857-4

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement, 41*(3), 178–194.

Collier, E. S., Normann, A., Harris, K. L., Oberrauter, L.-M., & Bergman, P. (2022). Making More Sustainable Food Choices One Meal at a Time: Psychological and Practical Aspects of Meat Reduction and Substitution. *Foods, 11*(9), Article 9. https://doi.org/10.3390/foods11091182

Collier, E. S., Oberrauter, L.-M., Normann, A., Norman, C., Svensson, M., Niimi, J., & Bergman, P. (2021). Identifying barriers to decreasing meat consumption and increasing acceptance of meat substitutes among Swedish consumers. *Appetite, 167*, 105643. https://doi.org/10.1016/j.appet.2021.105643

Graça, J., Godinho, C. A., & Truninger, M. (2019). Reducing meat consumption and following plant-based diets: Current evidence and future directions to inform integrated transitions. *Trends in Food Science & Technology, 91*, 380–390. https://doi.org/10.1016/j.tifs.2019.07.046

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes, 15*. https://doi.org/10.1186/s12955-017-0755-0

Ho, P. (2019). A new approach to measuring Overall Liking with the Many-Facet Rasch Model. *Food Quality and Preference, 74*, 100–111. https://doi.org/10.1016/j.foodqual.2019.01.015

Hobart, J. C., & Cano, S. J. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technology Assessment, 13*(12). https://doi.org/10.3310/hta13120

Jaeger, S. R., & Cardello, A. V. (2022). Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research. *Food Quality and Preference, 102*, 104676. https://doi.org/10.1016/j.foodqual.2022.104676

Johansson, M., Preuter, M., Karlsson, S., Möllerberg, M.-L., Svensson, H., & Melin, J. (2023). *Valid and Reliable? Basic and Expanded Recommendations for Psychometric Reporting and Quality Assessment.* OSF Preprints. Doi: 10.31219/osf.io/3htzc.

Lahne, J., Wolfson, J. A., & Trubek, A. (2017). Development of the Cooking and Food Provisioning Action Scale (CAFPAS): A new measurement tool for individual cooking practice. *Food Quality and Preference, 62*, 96–105. Doi: 10.1016/j.foodqual.2017.06.022.

Lai, J., Cella, D., Chang, C.-H., Bode, R. K., & Heinemann, A. W. (2003). Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Quality of Life Research, 12*(5), 485–501. https://doi.org/10.1023/A:1025014509626

Lee, Y.-W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q 3 statistics in an EFL reading comprehension test. *Language Testing - LANG TEST, 21*, 74–100. https://doi.org/10.1191/0265532204lt260oa

Leung, C. W., Wolfson, J. A., Lahne, J., Barry, M. R., Kasper, N., & Cohen, A. J. (2019). Associations between Food Security Status and Diet-Related Outcomes among Students at a Large, Public Midwestern University. *Journal of the Academy of Nutrition and Dietetics, 119*(10), 1623–1631. https://doi.org/10.1016/j.jand.2019.06.251

Mallinckrodt, B., & Tekie, Y. T. (2016). Item response theory analysis of Working Alliance Inventory, revised response format, and new Brief Alliance Inventory. *Psychotherapy Research, 26*(6), 694–718. https://doi.org/10.1080/10503307.2015.1061718

McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and STAN. *Routledge & CRC Press.* https://www.routledge.com/Statistical-Rethinking-A-Bayesian-Course-with-Examples-in-R-and-STAN/McElreath/p/book/9780367139919.

Melin, J., Fornazar, R., Spångfors, M., & Pendrill, L. (2020). Rasch analysis of the Patient Participation in Rehabilitation Questionnaire (PPRQ). *Journal of Evaluation in Clinical Practice, 26*(1), 248–255. https://doi.org/10.1111/jep.13134

Niimi, J., Sörensen, V., Mihnea, M., Valentin, D., Bergman, P., & Collier, E. S. (2022). Does cooking ability affect consumer perception and appreciation of plant-based protein in Bolognese sauces? *Food Quality and Preference, 99*, 104563. https://doi.org/10.1016/j.foodqual.2022.104563

Pendrill, L. R., & Fisher, W. P. (2015). Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. *Measurement, 71*, 46–55. https://doi.org/10.1016/j.measurement.2015.04.010

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value in Health, 18*(1), 25–34. https://doi.org/10.1016/j.jval.2014.10.005

Pickard, A. S., Dalal, M. R., & Bushnell, D. M. (2006). A Comparison of Depressive Symptoms in Stroke and Primary Care: Applying Rasch Models to Evaluate the Center for Epidemiologic Studies-Depression Scale. *Value in Health, 9*(1), 59–64. https://doi.org/10.1111/j.1524-4733.2006.00082.x

R Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* https://www.R-project.org/.

Rasch, G. (1960). *Studies in mathematical psychology: I.* Probabilistic models for some intelligence and attainment tests: Nielsen & Lydiche.

Regnault, A., Boroojerdi, B., Meunier, J., Bani, M., Morel, T., & Cano, S. (2019). Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? Learnings from the Parkinson's progression marker initiative cohort. *Journal of Neurology, 266*(8), 1927–1936. https://doi.org/10.1007/s00415-019-09348-3

Revelle, W. (2016). *How To: Use the psych package for Factor Analysis and data reduction.* Evanston, IL: Northwestern University, Department of Psychology.

Shimada, D., & Katahira, K. (2022). Sequential dependencies of responses in a questionnaire survey and their effects on the reliability and validity of measurement. *Behavior Research Methods.* https://doi.org/10.3758/s13428-022-01943-z

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*(1), 33. https://doi.org/10.1186/1471-2288-8-33

Stelmack, J., Szlyk, J. P., Stelmack, T., Babcock-Parziale, J., Demers-Turco, P., Williams, R. T., & Massof, R. W. (2004). Use of Rasch person-item map in exploratory data analysis: A clinical perspective. *Journal of Rehabilitation Research and Development, 41*(2), 233–241. https://doi.org/10.1682/jrrd.2004.02.0233

Tierney, N., & Cook, D. (2023). Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. *Journal of Statistical Software, 105*, 1–31. https://doi.org/10.18637/jss.v105.i07

Trubek, A. B., Carabello, M., Morgan, C., & Lahne, J. (2017). Empowered to cook: The crucial role of "food agency" in making meals. *Appetite, 116*, 297–305. https://doi.org/10.1016/j.appet.2017.05.017

Van de Winckel, A., Kozlowski, A. J., Johnston, M. V., Weaver, J., Grampurohit, N., Terhorst, L., … Mallinson, T. (2022). Reporting Guideline for RULER: Rasch Reporting Guideline for Rehabilitation Research: Explanation and Elaboration. *Archives of Physical Medicine and Rehabilitation, 103*(7), 1487–1498. https://doi.org/10.1016/j.apmr.2022.03.019

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

Wolfson, J. A., Lahne, J., Raj, M., Insolera, N., Lavelle, F., & Dean, M. (2020). Food Agency in the United States: Associations with Cooking Behavior and Dietary Intake. *Nutrients, 12*(3). https://doi.org/10.3390/nu12030877. Article 3.

Wright, & Masters, G. N. (1982). *Rating scale analysis.* Mesa Press.

Zagata, L., Urban, J., & Uhnak, T. (2022). Revalidation and empirical application of the Cooking and Food Provisioning Action Scale (CAFPAS). *Food Quality and Preference, 99*, 104540. https://doi.org/10.1016/j.foodqual.2022.104540