medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Systematic review of instruments for assessing culinary skills and other related concepts in adults: What is the quality of evidence of their psychometric properties?

Aline Rissatto Teixeira[1¶*]; Daniela Bicalho[1¶]; Betzabeth Slater[1¶]; Tacio de Mendonça Lima[2¶]

[1] Department of Nutrition and Public Health, School of Public Health, University of São Paulo, São Paulo, SP, Brazil.

[2] Department of Pharmaceutical Sciences, Federal Rural University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

* Corresponding author

E-mail: alinert@usp.br (ART)

¶These authors contributed equally to this work.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

Background: Culinary skills and food practices are important objects of study in the field of Public Health. Studies that propose to develop instruments for assessing such constructs show lack of methodological uniformity to provide evidence of validity and reliability of their instruments.

Objective: To identify studies that have developed instruments to measure culinary skills and other related concepts in adult population, and critically assess their psychometric properties.

Design: A systematic review was conducted. A literature search was performed in PubMed/Medline, Scopus, LILACS, and Web of Science databases until June 2019. The Directory of Open Access Journals and Google Scholar databases were searched to identify relevant grey literature. Searching, selecting and reporting were done according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement. Two reviewers were independently involved in study selection, data extraction, and instrument quality assessment. A third reviewer resolved all disagreements.

Results: The search identified 1428 potentially relevant studies, out of which 18 had potentially relevant records and 8 met the inclusion criteria. Studies used literature, experts' judgement, or qualitative interviews to develop the instruments. No studies received positive scores for all validity criteria. Although most studies received positive scores for internal consistency, none of them received positive scores for stability or presented evidence for content validity. One study showed positive results for construct validity. Two studies reported criterion validity, whose scores were deemed negative.

Conclusions: Many studies that surveyed culinary skills and related latent phenomena were identified. The overall quality of the psychometric properties of most instruments was considered insufficient, especially for validity measures. A universal definition of culinary skills as an overarching construct is recommended. The flaws observed in these studies show that there is a need for ongoing research in the area of the psychometric properties of instruments assessing these constructs.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

KEYWORDS: culinary skills; instruments; psychometrics; validity; reliability

Introduction

The discussion about the improvement of culinary skills and food practices has proven to be an important object of study in the field of Public Health; these skills are key factors associated with eating behaviors and with several complexities that represent social determinants of health [1].

Several authors define the term culinary skills in their publications [2-7], however, there is no a consensus on the definition of cooking skills or a consistent theoretical debate about it [4]. In summary, these skills are represented by a set of domains inherent in the practice of cooking [2-7], such as 1) knowledge, which includes nutritional and culinary knowledge (terms and techniques especially those considered healthy, involving the use of natural or minimally processed food and cooking from scratch) and sanitary hygiene control; 2) purchase planning, which concerns budget shopping, choice of ingredients, and organization of time for meal preparation; 3) creativity, which includes cooking meals with available ingredients and leftovers; 4) mechanical skills, which include the execution of slicing, cutting, heating, grilling, storing, and other cooking techniques; 5) food perception, which considers the ability to judge sensory perception of ingredients and their combinations; 6) confidence (self-efficacy), a dimension that might predict the cooking behavior at home; and 7) multi-tasking skills, which refer to the ability to perform different tasks simultaneously.

Culinary skills are associated with other concepts that involve the practice of proper and healthy eating, such as food literacy, which takes into account the broader social and environmental dimensions of eating together, associated with an individual's abilities [8]. Those considered to be "food literate" have the skills and abilities to revise and adapt their diet and food sources in response to changes imposed by modern life to maintain dietary quality [8]. Another concept related to culinary skills is food agency, which is related to the ability to act intentionally to change their own food environment. In general, its focus is on the individual mechanisms that lead to the act of cooking at home, secondary to other external elements that impact on the freedom of the individual

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and, consequently, on their autonomy [9]. Culinary autonomy is defined as the ability to think, decide, and act, to cook meals at home using mostly fresh and minimally processed foods, under the influence of interpersonal relationships, the environment, cultural values, access to opportunities, and the guarantee of rights; therefore, culinary skills represent an important dimension of this construct [10].

Time devoted to cooking has decreased and has been viewed as a global trend: food industry investments in advertising and marketing to "solve the everyday food problem" devalue cooking as an emancipatory competence associated with a healthy food routine [11]. Such decrease is associated with greater purchase of ultra-processed foods, and concerns public health experts around the world, considering their negative nutritional attributes and possible harmful effects on consumers' health, such as overweight, obesity, cancer and other chronic diseases and addiction-like behavior [12; 13]. It is worth mentioning that culinary practices, are also related to environmental, social and economic implications. Therefore, the valuing of the day-by-day cooking should be central in food and nutrition educational actions as an emancipatory and self-care practice [14].

The main source of cooking knowledge and skills is through parents [15; 16; 17; 18]. This information highlights the importance of adult cooking skills as a role model in food preparation habits development in children and young adults. In addition, Sidenvall *et al.* (2001) [19] found from a literature review that when changes in household dynamics happen (e.g., when a child moves away from the family or a divorce), the food provider may change their food habits and frequency of meal preparation, which may negatively affect their food choices.

In this scenario, culinary skills among adults, especially those responsible for preparing household meals, have been an important focus of research [15; 16; 20]. Among the publications on this subject are studies that propose to develop instruments that measure culinary skills and other related constructs (e.g., food literacy, food agency, food competency) in adults through the analysis of their psychometric properties.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Before being considered suitable, the instruments must offer accurate, valid, and interpretable data for the population's assessment. Moreover, the measures are supposed to provide scientifically robust results. These results are established based on measures of reliability and validity of the instruments [21-23]. Reliability is the ability to reproduce a consistent result in time and space or from different observers, demonstrating aspects of stability and internal consistency. It is one of the main quality criteria of an instrument [22]. Validity refers to the fact that a tool measures exactly what it proposes to measure, based on extent theory research and experts' judgement (content validity), the degree in which a group of variables really represents the construct to be measured (construct validity) and the degree in which the instrument is related to some external criterion, considered a widely accepted measure (criterion validity) [22-24].

There are public health policies focused on cooking in several parts of the world [4]. Despite the importance of developing instruments that measure culinary skills and related constructs as a strategy to assist the planning food and nutrition educational actions based on culinary practices, studies have shown lack of methodological uniformity to provide evidence of validity and reliability of their developed instruments. Moreover, other systematic reviews aimed to assess evidence of psychometric properties of instruments developed for different healthcare areas [25; 26]. However, so far we did not find any studies that propose the evaluation of psychometric properties of existing instruments that measure culinary skills and related concepts, which justifies the importance of this study, given the fact that the diagnosis of one's skills entrusted to the application of these instruments may be flawed, which could result in planning inappropriate food and nutrition educational actions for providing emancipatory and self-care practices.

Therefore, this systematic review aimed to identify studies that have developed instruments to measure culinary skills and other related concepts in adult population, and critically assess their psychometric properties.

We hope that this study can provide evidence-based guidance on the psychometric properties of instruments measuring culinary skills and related constructs, to subsidize the selection

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

of valid and reliable instruments by healthcare professionals to assess these subjects in clinical and public health settings and avoid unrealistic expectations about the information that such measures may provide.

Methods

The protocol of this systematic review was registered on the International Prospective Register of Systematic Reviews (PROSPERO database; http://www.crd.york.ac.uk/PROSPERO/; registration number CRD42019130836) and can be found in the S1 Appendix. The PRISMA [27] (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting systematic reviews were used to undertake the present review (S1 Table).

Search strategy

A comprehensive literature search for articles published until June 13, 2019, was performed in the Scopus, LILACS, PubMed, and Web of Science databases. The search strategy included the use of MeSH terms or text words related to the culinary skills, instruments, and validation studies. The PubMed/Medline search strategy was adapted from Terwee, Jansma, Riphagen *et al*. [28]. The full search strategy for all databases can be found in the S2 Appendix. In addition, a grey literature search was conducted in the Directory of Open Access Journals (DOAJ) (https://doaj.org/) and Google Scholar to identify studies not indexed in the databases listed above. Moreover, references to the articles found were also evaluated manually to include any potential studies that had not been identified.

Study selection

To be included in this review, the articles had to meet the following criteria: 1) being published in English, Portuguese or Spanish; 2) showing an original instrument; 3) describing a

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

literature search, combined or not with group discussions, to develop the instrument; 4) addressing culinary skills (defined as the skills related to confidence, practice, and knowledge to perform culinary tasks, from menu planning and purchasing food to combining ingredients and applying different culinary techniques, considering the daily routine and healthy eating), or other related concepts (food literacy, food agency, food autonomy, cooking confidence or self-efficacy; cooking competency); 5) describing the instrument validity studies.

Studies were excluded if: 1) they were applied to children and adolescents or used university students sample for analysis of psychometric properties of the instrument, considering that generalizing from students to the general adult public can be problematic when personal and attitudinal variables are used, as students vary mostly randomly from the general public [29; 30] or 2) if they were cross cultural adaptations of instruments, since they were not considered original instruments.

For the process of initial screening of abstracts and titles, we used the *Rayyan Web Platform for Systematic Reviews* [31]. Two authors (A.R.T. and D.B.) independently screened the titles and abstracts of citations to identify potentially relevant studies. Full-text articles were obtained and reviewed for further assessment according to the inclusion and exclusion criteria. When the full text could not be obtained, the corresponding authors were contacted by e-mail or other tools, such as ResearchGate (www.researchgate.net). All disagreements were resolved by the third author (T.M.L.).

Data extraction and analysis

Data extraction was performed independently by two authors (A.R.T. and D.B.) using a preformatted spreadsheet in Microsoft Excel. Disagreements were resolved by a third reviewer (T.M.L.).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The following information was collected: country and year of publication, participants, setting, sample size, format of instrument, target public, number of items of the instrument, development methodology, instrument domains, and the psychometric properties of the instrument.

Quality of psychometric properties

The psychometric quality of the instruments was determined according to the rating system adapted from Hair Jr, Black, Babin *et al*.[32]; Pedrosa, Suárez-Álvarez, and García-Cueto [33]; and Terwee *et al.* (2007) [34]. The criteria addressed the following properties: a) reliability, including internal consistency and stability; b) validity, including content, face, construct, and criterion. Each measurement property was reported to be positive (+), indeterminate (?), negative (-), or no information available (0), and properties are defined in Table 1. Two independent authors (A.R.T. and D.B.) applied this rating system, and any divergences between them were resolved by a third reviewer (T.M.L.).

Table 1. Quality criteria for psychometric properties of measurement (adapted from Hair Jr *et al.* [32], Pedrosa *et al.* [33], and Terwee *et al.* [34]).

PROPERTY	DEFINITION	RANKING**	QUALITY CRITERIA
Reliability Internal consistency	Extent to which items in a scale are intercorrelated (consistency among the variables)	+	Cronbach's alpha between 0.70 and 0.95
		?	Unclear design or method
		-	Cronbach's alpha <0.70 or >0.95, despite adequate design and method
		0	No information found on internal consistency
Stability* (reproducibility /test-retest)	Extent to which the instrument is stable over time, given by the agreement among individuals who are evaluated twice	+	ICC OR weighted Kappa ≥0.70
		?	Unclear design or method (e.g., time interval not mentioned)
		-	ICC OR weighted Kappa <0.70, despite adequate design and method
		0	No information found on stability

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Validity Content (including face validity*)	Extent to which the domain of interest is comprehensively sampled by the items in the instrument	+	Clear description is provided about the aims of the instrument, the target population, the concepts that are being measured, the item selection, AND the investigators OR experts that were involved in the items selection Quantitative evaluation: CVI = good level of statistical significance considering minimum values set at an associated probability of 0.05 or CVI value >0.8; CVR: minimum content validity ratios for item maintenance to ensure unlikely random agreement established at 0.05 significance and based on the number of evaluators or CVRc (critical CVR), considering CVR values higher than CVRc, established at 0.05 significance level and based on the number of evaluators for item maintenance
		?	Lack of clear description about the aspects mentioned above OR only the target population involved OR unclear method or design OR incomplete evaluation
		-	No target population involvement Quantitative evaluation: inappropriate level of statistical significance for CVI (>0.05) or CVI value < 0.8; inappropriate level of significance to ensure unlikely random agreement for CVR (>0.05) or CVRc (critical CVR)
		0	No information found on content validity
Construct	Extent to which a set of measured variables actually represents the theoretical latent construct those variables are designed to measure	+	Factor analysis performed on adequate sample size (minimum ratio of 5:1 and > 100) AND Bartlett's sphericity test (p < 0.005) OR KMO adequacy test (≥0.7) OR factors explaining ≥60% of the variance OR RMSEA ≤0.07 OR GFI and AGFI ≥0.95 OR SRMR ≤0,08 OR CFI ≥0.95
		?	No factor analysis OR unclear design or method
		-	Factor analysis performed on inadequate sample size AND/OR Bartlett's sphericity test (p >0.005) AND/OR KMO adequacy test (<0.7) AND/OR factors explaining <60% of the variance
		0	No information found on construct validity
Criterion	Extent to which the scores	+	Convincing arguments that gold standard is ‘‘gold’’ AND

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

	of an instrument relate to the scores of a gold standard measurement		correlation with gold standard ≥0.70
		?	No convincing arguments that gold standard is ''gold'' OR unclear design or method
		-	Correlation with gold standard <0.70, despite adequate design and method
		0	No information found on criterion validity

** +: positive rating; ?: indeterminate rating; -: negative rating; 0: no information available. Abbreviations: AGFI: Adjusted Goodness of Fit Index; CFI: Comparative Fit Index; GFI: Goodness of Fit Index; ICC: Intraclass Correlation; KMO: Kaiser Meyer-Olkin; RMSEA: Root Mean-Square Error of Approximation; SRMR: Standardized Root Mean-Square Residual.

Results

Search results

The electronic search (including gray literature databases) identified 1428 potentially relevant studies. After reviewing the titles and abstracts, eighteen articles were selected for full-text examination. Of these, eight studies [35-42] met the inclusion criteria and were included for review. A list of the excluded studies is shown in the S2 Table. No relevant studies were identified by searching the related articles and the reference lists of the included studies. A flowchart of the literature search is shown in Fig 1.

Fig 1. Study selection flowchart of literature search. Abbreviations: DOAJ: Directory of Open Access Journals; LILACS: Latin American and Caribbean Health Sciences Literature.

Characteristics of the studies

Studies were carried out in Brazil (1 study) [40], Denmark (1 study) [36], the United States of America (2 studies) [38;41], the United Kingdom (1 study) [35], Northern Ireland and Republic of Ireland (one study) [39], Netherlands (1 study) [42], and Australia (1 study) [37]. All of them were published in English. Most studies were published between 2017 and 2019 [36-40; 42]. One study did not seek ethical approval [35].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The studies were performed on different participants: parents of schoolchildren responsible for food preparation at home (one study) [40]; adult Danish consumers with variable household incomes (one study) [36]; adults living in the United States (US), but not necessarily US citizens (one study) [38]; adults from Northern Ireland and the Republic of Ireland who are responsible for preparing a main meal at least once per week (one study) [39]; men and women from low-to middle-income households (three studies) [35;37;41]; and adults, mostly highly educated, and dietitians (one study) [42]. Study samples were mostly composed of women [35; 37; 40-42]. Sample sizes ranged from 51 to 1049 individuals.

Some studies developed instruments to assess one domain related to culinary skills, such as cooking self-efficacy (confidence (two studies)) [35; 40], cooking competencies (experience and knowledge (one study)) [36], or multiple domains related to such phenomenon, in order to measure the effectiveness of a culinary and nutrition education program [41]. Other studies developed instruments aimed at evaluating latent phenomena related to culinary skills, such as food agency (one study) [38], food skills (two studies) [35; 39], and food literacy (two studies) [37; 42]. In these studies, culinary skills were established as one of the domains of the evaluated phenomenon or were presented in items belonging to one of the factors of the developed instrument.

All studies used literature combined with techniques such as focus groups, expert panels, and qualitative interviews to develop the instrument. The number of items ranged from ten to sixty-four. The instruments' domains were miscellaneous, ranged from one to eight and approached culinary skills by presenting items related to knowledge [35-37;41;42] (e.g., 'What is the term for preparing all ingredients, gathering equipment, and organizing your work area before beginning to cook?'[41] or 'Do you wash fruit and vegetables that don't need to be peeled before eating?' [35]); confidence [35;38;40;41] (e.g., 'How confident do you feel in cooking beans in pressure cooker' [40] or 'Indicate the extent to which you feel confident about cooking from basic ingredients' [41]); purchase planning and meal planning [37-39] (e.g., 'How long have you done the following action last month: plan meals ahead of time' [37] or 'On a scale of 1–7, where 1 is very poor and 7 is very

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

good how good are you at: plan how much food to buy' [39]); creativity [39; 41] (e.g., 'On a scale of 1–7, where 1 is very poor and 7 is very good how good are you at: cook a healthy meal with only few ingredients on hand' [39] or 'During the past month how often did you reuse leftovers for another meal' [41]); food perception [39;40;42] (e.g., 'On a scale of 1–7, where 1 is very poor and 7 is very good how good are you at: use herbs and spices to flavor dishes'[39] or 'Are you able to see, smell or feel the quality of fresh foods?' [42]), mechanical skills [35-39; 41; 42] (e.g., 'Are you able to prepare fresh fish in different ways?' [42] or 'On a scale of 1–7, where 1 is very poor and 7 is very good how good are you at: steam food, or chop vegetables, or cube meat' [39]); and multi-tasking skills [38] (e.g., 'My family responsibilities prevent me from having time to prepare meals' [38], or 'My social responsibilities prevent me from having the time to prepare meals' [38]).

The studies reported analysis of the psychometric properties of their instruments: Six studies reported internal consistency, face validity, literature review, or experts' judgment for content validity and construct validity [36-39; 41; 42]. No studies presented quantitative evaluation for content validity. Two studies did not report construct validity, and used stability for analysis [35; 40]. Two studies [38; 42] reported criterion validity. The characteristics of the included studies are shown in Table 2.

Table 2. Descriptive data and characteristics of the included studies.

Authors (years), country	Participants/ Setting; sample size[a]	Latent phenomenon evaluated	Format/ Target population	No. of items	Development	Domains evaluated (number/name)	Psychometric properties
Martins, et al. (2019), Brazil	Parents of schoolchildren responsible for food preparation at home/ n=51	Cooking skills self-efficacy	Index/ Adult population	10	2 evaluations with experts in nutrition and public health belonging to the research group that supported the preparation of the DGBP[b]	(1)/ Cooking skills self-efficacy (items related to confidence)	*Content validity (including face validity):* experts in nutrition and public health belonging to the research group that supported the preparation of the DGBP[b], and pilot study (10 parents of schoolchildren) *Reliability and internal consistency*: Cronbach's alpha=0.75 *Stability:* Test-retest at an interval of seven to 15 days: Moderate reproducibility by quadratic weighted Kappa (0.55)
Bech-Larsen & Tsalis (2018), Denmark	Adults (18 years or older) Danish consumers/ 2 surveys (n1=418; n2=1008)	Cooking competency	Cumulative scales Adult population	*Survey 1:* 30 *Survey 2*: 35	1 focus group interview and 5 individual interviews with specialists	*Survey 1:* (2)/ Experience; Knowledge (items related to mechanical skills and knowledge) *Survey 2:* (3)/ Experience; Knowledge; Satisfaction with food-related lifestyle	*Content validity:* 1 focus group interview and 5 individual interviews *Survey 1 (construction):* consultation with expert sources (two nutritionists). Coe□cients of reproducibility (CR) and scalability (CS) calculations and progressive elimination of the item with most respondents who o□ered disconfirmed progressions. RC (>0.85) and CS (>0.6) *Survey 2 (nomological validity and reliability):* partial least squares (PLS) applied in a second-order structural equation model: respondents' cooking experience and knowledge account for 10.5% of the variance Convergent validity: composite reliability indices >0.8, and average variance extracted for each variable: > 0.5.
Lahne et al. (2017), USA	Adults (18 years or older)/ Development sample: n= 445; Validation sample: n= 500	Food agency	Scale/ Adult population	28	Literature review and expert panels	(3)/ Self-Efficacy (items related to confidence, mechanical skills and one item related to food perception, regarding the use of herbs and spices); Attitude; Structure (items related to purchase planning and meal planning, mechanical and multi-tasking skills)	*Content validity:* expert panels (community nutritionists, chefs, food scientists, extension officers, and rural sociologists). *Construct validity*: Maximum-Likelihood EFA with 3 factors and oblique (promax) rotation using the Full Sample data and Structural Equation Modeling (SEM - CFA) to evaluate structural stability and generalizability Variance explained: 47%; Root Mean Squared Residuals (RMSR): 0.03. *Reliability:* Cronbach's alpha > 0.70 *Criterion Validity*: correlation between FIS [43] [c] and the CAFPAS[d] (r= 0.65 [-0.61, 0.69]).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Lavelle et al. (2017), Island of Ireland (Northern Ireland and Republic of Ireland) (Study 1)	Adults (20 - 60 years old) responsible for preparing a main meal at least once per week/ n=1049	Cooking skills and food skills	Scale/ Adult population	33	Literature review and expert opinion	(2)/ Cooking skills (items related to mechanical skills an 1 item relate to food perception); Food skills (items related to purchase planning and meal planning, creativity)	*Content validity (including face validity):* 4 qualitative interviews with experts working in the area of health promotion including cooking and food skills interventions and education and literature review Pilot testing: 40 internal pilot surveys and 14 pilot survey field based interviews to assess clarity of the questions and how easy participants found the measures to complete *Construct validity:* EFA using oblique rotation (direct oblimin). 3 factors with eigenvalues > 1 accounting for 65.26%, 8.2%, and 3.2% of the variance. All factor loadings above the minimum criterion *Reliability (internal consistency):* Cronbach's alpha for the cooking skills confidence=0.93; for the food skills confidence = 0.94
Begley et al. (2018), Australia	Australian adults from low to middle-income households, participants in the Western Australia nutrition and cooking FSA program, developed by the WA Foodbank/ n=882	Food literacy	Questionnaire Adult population attending the FSA program	14	Literature review. Feedback on content and format of the questionnaire by Foodbank program facilitators. Discussion with food literacy experts; Observation and discussion with participants for feedback on the wording of questions and relevance of some of the food literacy behaviors. Discussions with the Department of Health WA	(3)/ Plan & Manage; Selection; Preparation (items related to purchase planning, meal planning, knowledge and mechanical skills)	*Content validity (including face validity)*: Empirical research on the domains of food literacy. Feedback on content and format of the questionnaire by Foodbank program facilitators. Discussion with food literacy experts and the Department of Health's Nutrition Monitoring Surveillance Survey Observation and discussion with participants for feedback on the wording of questions and relevance of some of the food literacy behaviors *Construct Validity:* EFA: adequate sample size was indicated by KMO of 0.859. Bartlett's test of sphericity ($p < 0.0001$); Scree plot = 3 factors: Two items, loaded on 2 factors. 3 items that did not meet the specified loading for any factor were maintained, but were not included in the EFA and Cronbach's alpha analysis *Reliability (internal consistency):* Cronbach's alpha = 0.79, 0.76, and 0.81 for each factor: Plan & Manage, Selection, and Preparation, respectively
Barton et al. (2011), UK - Scotland	Adults of low income communities, mostly female, participants in	Cooking confidence and food skills	Questionnaire Adult population of low income	19	Literature review and assessment of the original Cook Well programme evaluation tools.	(5)/ Meal preparation; Confidence in cooking and tasting; Usual food consumption patterns;	*Content validity*: Assessed by a panel of public health experts and face validity by 20 adults who were not involved in cooking skills intervention classes but were typical of individuals who attend community classes *Stability:* Repeat reliability calculated by Spearman correlation

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

	cooking interventions (Cook Well programme)/ Time 1 (face validity and internal consistency): n= 74 Time 2 (test-retest): n=57; feasibility testing: n=13		communities		Panel of public health experts	Knowledge about fruit and vegetables; Knowledge of good practice (items related to mechanical skills, confidence and knowledge)	coefficients (57 respondent adults attending community-based classes other than cooking): for each item = range 0.46–0.91 *Internal consistency:* Cronbach's alphas = 0.86 and 0.84 for the confidence and knowledge questions, respectively Ethical approval was not sought for the study
Poelman et al. (2018), Netherlands	Dutch adults recruited via the Facebook page and Twitter account of The Netherlands Nutrition Centre (n= 755) and dieticians registered as members of the Netherlands Association of Dieticians (NAD) (n=207)	Food literacy	Scale/ Adult population	29	Experts' insights and literature review	(8)/ Food preparation skills (items related to mechanical skills and one item related to food perception); Resilience and resistance; Healthy snack styles; Social and conscious eating; Examining Food Labels; Daily food planning; Healthy budgeting; Healthy food stockpiling (items related to nutrition knowledge)	*Content validity*: expert meeting with health professionals and academics working in the field of food literacy (n=10) and an additional group of Dutch experts (academics, dieticians, and health professionals) in the field of food consumption and health literacy (n=17) *Construct validity:* PCA (Principal Component Analysis). KMO measure of sampling adequacy = 0.88 and Bartlett's Test was statistically significant (p<.000). None of the items showed a correlation > 0.90. PCA resulted in 12 components with an eigenvalue > 1.0. The scree plot indicated eight rather than 12 components to be sufficient. PCA with eight components after removing items that did not show component-loadings or loaded on multiple components showed explained variance = 62% *Internal consistency:* Cronbach's α=0.83. One subscale had a lower than sufficient (0.7) Cronbach's alpha of 0.58, indicating that this subscale in isolation showed inadequate reliability *Convergent and discriminant validity:* Self-control and food literacy positively correlated to a moderate magnitude (r=0.51, p =<.001); impulsiveness and food literacy negatively correlated (r=−0.31, p=<.01. *Criterion validity:* self-perceived food literacy was positively associated with healthy food consumption and negatively with unhealthy food consumption.
Michaud (2007), USA	Adult parents and caregivers in South Carolina	Cooking skills and behavior	1 index, 6 scales, and 1 evaluation tool/	64	Questions were determined by the CWC program goals, literature,	(8)/ Availability and Accessibility of Fruits and Vegetables (AAFV); Cooking	*Content validity* by experts' review. *Pilot study:* test-retest reliability: correlation coefficient results ranged from r=.43 (SEFVS) to r = .89 (SECT). *Test-retest reliability* low for the CB (r=.48) and SEFVS (r = .43)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

	recruited from church preschools, Head Start preschools, and public elementary schools and playgroup participants in a culinary nutrition education program/ Pilot reliability (n = 39); EFA (n-162)	Adult population participants in the Cooking With a Chef culinary nutrition education program	and experts' review	Attitude (CA); Cooking Behavior (CB); Self-Efficacy in Produce Consumption (SEPC); Self-Efficacy in Cooking (SEC); Self-Efficacy in Using Basic Cooking Techniques (SECT); Self-Efficacy in Using Fruits, Vegetables, and Seasonings (SEFVS); Knowledge of Cooking Terms and Techniques (CTT) (Items related to knowledge, confidence, mechanical skills and creativity)	scales. *Internal consistency* (Cronbach's alpha): ranged from .15 for the CB scale to .89 for the CA and SECT scales. *Construct validity*: EFA with promax rotation performed on the six scales and the AAFV index. Scree plot eigenvalues used to confirm the number of factors on a scale. Cronbach's alpha was computed for each scale ● AAFV index - All items retained on a single loading at least 0.30 except item four (factor loading =0.24). Cronbach's alpha= 0.51 (item four was retained) ● CA scale - seven items on two factors: Positive Attitude subscale (n = 3) and Negative Attitude subscale (n = 4). Cronbach's alpha = 0.79, all items loadings at least 0.30 ● CB scale - All items were retained on one factor. Factor loadings < 0.30 for the first two items and 0.38 for the third item. Cronbach's alpha=0.15 ● SEPC scale - all items retained on a single factor. Factor loadings ranged from 0.63 to 0.74. Cronbach's alpha = 0.78 ● SEC scale - six items retained on a single factor. Factor loadings ranged from 0.57 to 0.69. Cronbach's alpha = 0.79 ● SECT scale - the twelve initial items retained on a single factor. Factor loadings ranged from 0.39 to 0.77. Cronbach's alpha = 0.87 ● SEFVS scale - all four original items retained on a single factor. Items ranged from 0.62 to 0.80. Cronbach's alpha = 0.80. Knowledge of Cooking Terms and Techniques Evaluation: High knowledge scorers (n = 89): 75% of correct answers = higher scores on the evaluation. Low knowledge scorers = 62.5% of correct answers on the evaluation. Correlations between all scales and indices were computed for the entire larger study population and separately for low and high knowledge scores: = r <0,7

[(a)]Sample size for psychometric properties; [(b)] Dietary guidelines for the Brazilian population; [(c)] Food Involvement Scale; [(d)] Cooking and Food Provisioning Action Scale.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Quality of the psychometric properties

The corresponding results of the psychometric properties are shown in Table 3. Five studies [37-40; 42] obtained positive results and two studies received negative results [35; 41] for internal consistency. One study [36] did not determine Cronbach's alpha. Five studies reported stability [35; 36; 40-42]; however, none of these studies received positive rating because they presented results inferior to the minimum criterion for weighted Kappa despite adequate design and method [40; 41] or because they reported unclear methods [35;36;42]. All studies received indeterminate ratings for content validity. The authors did not calculate any index of agreement for content validity.

Table 3. Evaluation of quality criteria of studies on psychometric properties.

Authors (year)	Reliability		Validity		
	Internal consistency	Stability	Content	Construct	Criterion
Martins et al. (2019)	+	-	?	0	0
Bech-Larsen & Tsalis (2018)	?	?	?	?	0
Lahne et al. (2017)	+	0	?	-	-
Lavelle et al. (2017) (Study 1)	+	0	?	+	0
Begley et al. (2018)	+	0	?	-	0
Barton et al. (2011)	-	?	?	0	0
Poelman et al. (2018)	+	?	?	?	-
Michaud (2007)	-	-	?	-	0

Rating: (+) = positive; (?) = indeterminate; (-) = negative; (0) = no information available.

For information regarding the definitions of psychometric properties, see Table 1.

A positive rating was given to one study [39] for construct validity, considering sufficient sample size for Exploratory Factor Analysis (EFA) and factors that explain ≥60% of the variance and use of the oblique method (oblimin) as a rotational method. In addition, two studies [36; 42] received an indeterminate rating because they reported different statistical measures for construct validity. One of these studies used PCA instead of factor analysis [42]. Another study assessed

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

nomological validity [36]. Three studies [37; 38; 41] received a negative rating because they presented inadequate sample size for EFA [41] or factors that explain less than 60% of the variance [38] or because the retention of items did not meet the specified loading for any factor [37]. Most studies [35-37; 39-41] did not provide information on criterion validity. Studies [38; 42] that reported criterion validity did not describe it clearly (convincing arguments for gold standard), and showed correlation with gold standard < 0.70, therefore they received a negative rating.

Discussion

Summary of evidence

To our knowledge, this is the first systematic review to identify and appraise the studies that developed instruments for assessing culinary skills or related latent phenomena. This article has provided a comprehensive critical analysis of the studies' characteristics and their psychometric properties of measurement. Eight studies that developed instruments to measure and evaluate culinary skills and related phenomena were found.

This systematic review has highlighted gaps in these instruments, suggesting the need to develop new studies with robust and standardized psychometric methodology that shows validity and reliability. Although most studies received positive scores for reliability criteria, that is, internal consistency, none of the included studies received positive scores for stability. No studies received positive scores for all validity criteria, and none of them presented satisfactory evidence for content validity since the authors did not calculate any index of agreement. Only one study showed positive results for construct validity and two studies reported criterion validity, but their scores were deemed negative. These results indicate that while there are isolated measures that were appraised in this review that show good promise in terms of quality of evidence of psychometric properties, no studies presented satisfactory results for every aspects of reliability and validity.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

General view of the studies

The majority of the included studies presented items of the instruments to assess cooking knowledge, confidence and mechanical skills. Although these are important domains of culinary skills and related constructs involving culinary practices, these domains themselves do not guarantee the preparation of meals from basic ingredients. Many people lack the ideas (creativity), menu-planning skills or ability to judge flavor, color and texture of the combinations of ingredients, and the ability to multi-task within a demanding family lifestyle necessary to organize and prepare a homemade meal [7;20]. According to Ternier (2010) [15], when there are time constraints in a fast paced or stressful lifestyle, being able to do multiple tasks simultaneously is an advantage. Also, if the meal provider is unable to plan and organize a meal, or unable to create a meal that will satisfy those who are eating it, he or she may find it easier to buy a convenience food product that will save time and energy, and be satisfying to everyone [15].

Professionals involved with health promotion, should include cooking themes, in their meetings, presentations and discussions with the public [6]. Hence, subsidizing the choice of instruments that enable the assessment of culinary skills and healthy culinary practices, based on the aforementioned domains, is essential for Public Health scenario.

All studies presented their instruments in English. Although studies are mostly from countries whose native language is English, one study [40] developed an instrument for application with Brazilian parents of schoolchildren responsible for food preparation at home. Despite the authors' intention to provide access to their study through the use of universal language, translating the instrument to English is not enough to guarantee its international applicability, considering cultural aspects. Developing a new instrument in one's own language or adapting existing instruments to each setting is necessary to guarantee the instruments' linguistic and cultural appropriateness [44].

Most studies reported the development of scales, indexes, and questionnaires. One study classified their instrument as an index [40]; however, the instrument used Likert scale to register

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

participants' statements related to the assessed latent phenomenon (cooking self-efficacy). It is important to highlight differences between an index and a scale. An index compiles one score from an aggregation of two or more indicators that attempt to signal, by means of a value, both a content relation with the represented phenomenon and the evolution of a quantity in relation to a reference. The indicator communicates or reveals progress toward a certain goal, and it is applied as a resource to make a tendency or phenomenon not immediately detectable by isolated data more noticeable. It represents an essential tool for the decision-making process and social control, and it is not an explanatory or descriptive element, but provides punctual information on time and space, whose integration and evolution can activate or accompany reality [45].

A scale, on the other hand, measures levels of intensity at the variable level, like to what extent a person agrees or disagrees with a particular statement. A scale is a type of measure composed of several items that have a logical or empirical structure among them. The most commonly used scale is the Likert scale. The sum of scores for each of the statements creates an overall score of the intensity related to the assessed latent phenomenon [21].

Another study [36] reported the construct and validation of a set of cumulative scales to measure consumers' cooking knowledge and experience as well as the links with consumers' food-related life satisfaction. A Guttman (cumulative) scale consists of a number of items that are empirical indicators of some single variables or attitude continuum. In this discussion, the ordered response categories for all items are dichotomized, so that all responses are scored as positive or negative for the variable. The items can be ordered from high to low according to the proportion of persons scored as positive [46]. Examples of dichotomized items for the cumulative scales used in the included study [36] are as follows: for the developed experience scales, respondents indicated yes/no to a stem that began, "Did you, within the previous year, prepare….?"; for the knowledge scales, respondents indicated true/false in response to a stem that began, "Are the following statements…?"

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

To develop survey measurement instruments that attain the true responses from the population, one of the challenges is to form questions that not only capture the theoretical concept under evaluation, but also minimize the impact of the design characteristics on the quality of the responses. Although dichotomous scales require fewer interpretative efforts (which can harm consistency compared to rating scales), increasing the number of scale points appears to produce more valid measurements than forcing respondents to choose between two response categories [47].

Regarding the need for submission of psychometric studies for ethical approval, one study [35] justified the absence of ethical approval because it comprised developmental work for service evaluation. It is important to emphasize that, despite the fact that validation studies aim at the development of tools for measuring latent phenomena, methods applied to evidence the reliability and validity of such instruments involve the participation of human beings; therefore, the submission of such studies to ethical approval is not only essential, but also indispensable [48;49].

Psychometric quality

Although all instruments reported some psychometric information, the evaluation of the psychometric quality using the criteria adopted in this systematic review exhibited some missing data.

Regarding the reliability of the instruments, most studies reported internal consistency [37-40; 42]. Internal consistency is a measurement of the extent to which individual items of the instrument are correlated and produce consistent results of a concept or construct, through Cronbach's alpha coefficient [32]. Two studies obtained negative scores for reliability [35; 41]. One of them [35] tested two out of five sections of the questionnaire (related to confidence and knowledge) for internal consistency, and the other sections were not tested based on the justification that the domains within each section of the instrument assessed different constructs. The other study that received negative scores [41] showed adequate results for internal consistency, considering the overall scale, but two out of the eight scales developed in this study presented unacceptable values

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

for Cronbach's alpha. In addition, one item that showed low factor loading in one of these scales was retained. The authors' justification was that Cronbach's alpha would not reach the 0.7 acceptability level regardless of the item's removal. It is important to consider that Cronbach's alpha gives a unique value for any set of data and gives a value for the mean of the distribution of all possible coefficients of the parts that make up the instrument; moreover, this depends not only on the magnitude of the correlation between the items, but also on the number of items in the scale [23]. Therefore, the fewer the items removed from an instrument are, the less affected the alpha value will be.

Another study showed an internal consistency coefficient different from the criteria of this review and obtained an indeterminate score [36]. The authors chose to use coefficients of reproducibility and scalability, proposed, respectively, by Guttman [50] and Menzel [51] to analyze the reliability criterion of their cumulative scale to measure consumers' cooking knowledge and experience. The coefficient of reproducibility, proposed by Guttman [50], measures the extent to which an observed set of response patterns agrees with that expected from a perfect scale. A high value indicates close agreement and a value equal to or greater than 0.90 is usually seen as an indication of the existence of a scale. Criticism has been leveled at the use of this coefficient on the grounds that it can be expected to attain a high value even when the scale items are independent of one another. The expected value of the coefficient of reproducibility will vary according to both the number of items comprising the scale and the probability of a positive response to each item [52]. Increasing the number of items also decreases the coefficient of reproducibility. For example, for a five-item scale with probability of positive response within the range of 0-1, as shown in the included study, the maximum value of the coefficient of reproducibility is obtained, while for a seven-item scale with the same range of positive responses, lower results are expected. One approach to solving the problem of spuriously high reproducibility is the coefficient of scalability (CS) suggested by Menzel [51]. The CS measure has also been criticized on the grounds that it deals with the reproducibility of the items (or scores) rather than with the reproducibility of the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

scale. Thus, studies that rely on coefficients of reproducibility and scalability to show reliability may provide compromised results [52].

Three studies received indeterminate score for stability, considering the reason above [35; 36; 42]. Stability was assessed by test-retest in two studies [40; 41] that obtained negative scores. It is important to emphasize that although the test-retest is considered a criterion of stability by Terwee *et al.* [34], it is an association measure (correlation) intended to test the repeat reliability of the instrument, and it does not measure concordance, but the force of the relationship between variables. That is, the results show the consistency in responses between tests, not the accuracy of the instrument [24; 53].

Studies that only used the reliability criterion for analysis without other adequate criteria for the psychometric measurements of the instruments may not provide trustworthy results, because these instruments reproduce only a consistent result in time and space from different observer (reliability), without measuring exactly what they propose (validity) [53;54]. Two studies included in this review fit into this scenario [35; 40], since authors exclusively assessed the internal consistency (using Cronbach's alpha) and stability (using test-retest or another coefficient different from the quality criteria established in this review), as well as inappropriate content validity (disregarding existing empirical methods to quantify the degree of experts' agreement). Moreover, no construct or criterion validity tests were presented.

All studies included in this review failed to show proper content validity: most studies relied on face validity, literature research, and experts' judgment, however no index for content validity was calculated to confirm experts' judgment agreement, which can be considered a problem [33].

Face validity is the suitability of the content of a test or item(s) for an intended purpose as perceived by test takers, users, and/or the general public and represents a controversial form of perception based evidence to affirm if the test measures what it purports to measure [25]. Perception, however, is an interpretive process influencing each individual according to their experiences, knowledge, beliefs, and attitudes among other factors; therefore, it is generally agreed

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

that face validity may not represent sufficient evidence to support the interpretation and use of test scores [25; 55].

Content validity based on a quantitative approach, regarding the use of statistical methods derived from the experts' judgment, proves itself to be essential. Otherwise, the mere fact that the experts report on the lack or excess of items representative of the construct, or that they simply determine to what extent each element corresponds to the latent phenomena, does not itself provide relevant information for the validation process [23; 32; 33]. For this reason, it is essential to apply some of the existing empirical methods to quantify this degree of agreement [33; 56].

One study received positive score for construct validity [39], considering adequate sample size for factor analysis, and used statistical analysis with adequate percentage of variance. Construct validity refers to the extent to which a set of measured variables actually represents the theoretical latent construct those variables are designed to measure [23]. Evidence of construct validity provides confidence that item measures taken from a sample represent the actual true score that exists in the population and can be examined using factor analysis and multivariate regression models [32].

Three studies [37; 38; 41] obtained scores deemed negative for construct validity: one study [41] presented inadequate sample size to perform factor analysis and another study [37] reported that three items did not meet the specified loading for any factor and, still, were not removed. The third study [38] had factor analysis correctly performed; however, it showed unsupported results for adequate percentage of explained variance. The purpose of such criterion is to ensure practical significance for the derived factors by ensuring that they explain at least a specified quantity of variance, and 60% of the total variance is considered satisfactory [32].

Three [37;39;42] out of the six studies that reported construct validity [36-39;41;42] described the Kaiser-Meyer-Olkin (KMO) adequacy test with values > 0.80, which is considered very good for factor analysis appropriateness [32].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Two studies [36;42] received an indeterminate score for construct validity because they described different statistical models from the established criteria in this review: one study used nomological validity, which is one of the forms of validity that pertains to the testing of proposed relations among constructs in models, and considered a theoretical plausibility test proposed by Cronbach and Meehl in 1955 [57]. The researcher must identify theoretically supported relationships from prior research or accepted principles and then assess whether the scale has corresponding relationships [32]. That is, in order to provide evidence that the proposed measure has construct validity, it is necessary to develop a nomological network for such measure. This network would include the theoretical framework for what the researcher is trying to measure, an empirical framework for how the researcher proposes to measure it, and a specification of the linkages between these two frameworks. In other words, it provides the opportunity to specify patterns of relations among constructs that reflect mechanisms such as additive (multiple factors explain unique variance in outcomes), mediation (one or more factors serve to explain or transmit the effect of one variable on another), and moderation (one or more factors change the pattern of the effect of one factor on another) effects, thus reflecting the researcher's expectations as to how the phenomenon works [58].

Another study [42] used principal component analysis (PCA) instead of factorial analysis (FA) for construct validity. Both FA and PCA are techniques that aim to reduce a certain number of items to a smaller number of variables. Although there is a significant conceptual difference between these two data reduction techniques, they are generally used indiscriminately, impairing the interpretation and validity of results [59]. Factor analysis is used to estimate the unknown structure of the data. This is a critical point that distinguishes FA from PCA [60]. The last technique aims to describe a large dataset in a simpler dimension. In this case, PCA is used mainly to show graphically the relationships among the variables in some reduced dimension graphs. On the other hand, FA is a statistical model used to build patterns (factors), which are latent variables, to predict a phenomenon: it assumes a characteristic of the multivariate model by calculating factor loadings

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and errors assigned to each factor. One of the main differences between PCA and FA in mathematical terms is the values found in the diagonal of the correlation matrix: The total variance of each variable is a result of the sum of the shared variance with another variable, the common variance (communality), and the unique variance inherent in each variable (specific variance). In PCA, all variances are taken into account in the calculations, while in FA, only the common variance is used; therefore, the diagonal of the correlation matrix includes only communalities [23]. Hence, PCA does not provide valid substitute for factor analysis, since FA is a more complex method in the sense that factors reflect the causes of observed variables [61].

Regarding criterion validity, little information was available in the included studies. Only two studies [38; 42] presented criterion validity, but did not describe it clearly and did not obtain a satisfactory correlation with the gold standard. Lack of data also was reported in research in other areas that use similar criteria. These findings were expected since most of the time, the criterion validity is a challenge for the researcher, because it demands a “gold standard” measure to be compared with the chosen instrument, which cannot be easily found in all knowledge areas [22; 62].

Limitations

This review has some limitations. It is possible that some studies were missed out because they were not indexed in the databases searched, or were published for institutions, foundations, or societies. Information published in languages other than English, Portuguese or Spanish were not included; therefore, some research findings may have been overlooked. In addition, although the criteria were adapted from previous studies, the difficulty of interpreting the studies may have under- or overestimated the quality of the instruments' psychometric properties.

Conclusion

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

In this review, many studies that surveyed culinary skills and related latent phenomena were identified. Regarding the quality of evidence of psychometric properties, most instruments identified in these studies were considered insufficient, especially for validity measures. Thus, the flaws observed in these studies show that there is a need for ongoing research in the area of the psychometric properties of instruments assessing culinary skills or other related constructs. Moreover, our findings contribute to supporting the selection of valid and reliable instruments by healthcare professionals in clinical and Public Health settings.

In other hand, measuring culinary skills involves several separate but related domains, which integrate other constructs related to the culinary practices. Therefore, it is recommended that a more consistent and consensual definition of culinary skills as a construct be generated. Instruments should cover items and domains considering skills and not just knowledge, cutting and cooking techniques, and confidence in preparing meals in order to allow a greater understanding of barriers and facilitators related to the culinary practice.

Acknowledgements

We thank the team of librarians of the School of Public Health (University of Sao Paulo) for the specialized support in electronic databases and the research group of the Department of Nutrition and Public Health of the School of Public Health (University of Sao Paulo) for proof reading the article.

References

1. Lavelle F, McGowan L, Spence M et al. Barriers and facilitators to cooking from 'scratch' using basic or raw ingredients: A qualitative interview study. Appetite. 2016; 107, 383-91. doi: 10.1016/j.appet.2016.08.115

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

2. Foley W, Spurr S, Lenoy L. De Jong M, Fichera, R.. Cooking skills are important competencies for promoting healthy eating in an urban Indigenous health service. Nutrition & Dietetics. 2011 68 (4), 291-6. doi: 10.1111/j.1747-0080.2011.01551.x

3. Hartmann C, Dohle S, Siegrist M.Importance of cooking skills for balanced food choices. Appetite. 2013; 65, 125-31. doi: 10.1016/j.appet.2013.01.016

4. Jomori MM, Vasconcelos FDAGD, Bernardo GL, Uggioni PL, Proença RPC. The concept of cooking skills: A review with contributions to the scientific debate. Rev Nutr. 2018; 31 (1), 119-135. doi: 10.1590/1678-98652018000100010

5. Metcalfe J, Fiese B, Liu R, Emberton E, McCaffrey J. Innovative approaches to the evaluation of hands-on cooking skills with youth. J Nutr Educ Behav. 2018; 50(7), Suppl 7, S6. doi: 10.1016/j.jneb.2018.04.026

6. Melo EA, Jaime PC, Monteiro CA.Dietary Guidelines for the Brazilian population. 150 p. Brasília: Ministry of Health of Brazil. Secretariat of Health Care. Primary Health Care Department; 2015.

7. Short F.Kitchen Secrets: The meaning of cooking in everyday life. Oxford, UK: Berg Publishers; 2006.

8. Cullen T, Hatch J, Martin W, Higgins J, Sheppard R. Food literacy: definition and framework for action (Perspectives in practice/Perspectives pour la pratique). Can J Diet Pract Res. 2015; 76(3), 140-145. doi: 10.3148/cjdpr-2015-010

9. De Oliveira MFB. Autonomia culinária: desenvolvimento de um novo conceito. PhD Thesis. State University of Rio de Janeiro (UERJ); 2018.

10. Trubek AB, Carabello M, Morgan C, Lahne J. Empowered to cook: The crucial role of 'food agency' in making meals. Appetite. 2017; 116, 297-305. doi: 10.1016/j.appet.2017.05.017

11. Van Der Horst K, Brunner TA, Siegrist M. Ready-meal consumption: Associations with weight status and cooking skills. Public Health Nutr. 2011; 14(2), 239-45. doi: 10.1017/S1368980010002624

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

12. Aranceta J.Community nutrition. Eur J Clin Nutr. 2003; 57, Suppl 1, S79-S81. doi: 10.1038/sj.ejcn.1601823

13. Ludwig DS. Technology, diet, and the burden of chronic disease. JAMA. 2011; 305(13), 1352-1353. doi: 10.1001/jama.2011.380

14. Castro IRR de. Challenges and perspectives for the promotion of adequate and healthy food in Brazil. Cad. Saúde Pública. 2015; 31(1), 07-09. doi: 10.1590/0102-311XPE010115

15. Ternier S. Understanding and measuring cooking skills and knowledge as factors influencing convenience food purchases and consumption. SURG Journal. 2010; 3(2):69-76.

16. Bowen, R. L., & Devine, C. M. "Watching a person who knows how to cook, you'll learn a lot". Linked lives, cultural transmission, and the food choices of Puerto Rican girls. Appetite. 2011; 56(2), 290-298.

17. Nor NM, Sharif MSM, Zahari MSM, Salleh HM, Isha N, Muhammad R. The transmission modes of Malay traditional food knowledge within generations. Procedia-Social and Behavioral Sciences, 2012; 50, 79–88.

18. De Backer CJS. Family meal traditions. Comparing reported childhood food habits to current food habits among university students. Appetite. 2013; 69, 64 - 70. doi:10.1016/j.appet.2013.05.013

19. Sindevall B, Margaretha N, Fjellström C. Managing food shopping and cooking: the experiences of older Swedish women | Ageing & Society | Cambridge Core. Cambridge University Press. 2001.

20. Caraher M, Dixon P, Lang T, Carr-Hill R. The state of cooking in England: the relationship of cooking skills to food choice. Br Food J. 1999; 101(8), 590-609.

21. DeVellis RF.Scale Development. Theory and Applications. Chapel Hill, USA: SAGE Publications.2017.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

22. De Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. Epidemiol Serv Saude. 2017; 26, 649-659 doi: 10.5123/s1679-49742017000300022

23. Furr, RM, & Bacharach, VR.Psychometrics An introduction. 2nd ed.. London: Sage Publications.2014.

24. Echevarría-Guanilo ME, Gonçalves N, Romanoski PJ.Psychometric properties of measurement instruments: conceptual bases and evaluation methods - part I. Texto Contexto Enferm. 2018; 26(4): e1600017. doi: 10.1590/0104-07072017001600017

25. Brasil V, Oliveira G, Moraes KL.Psychometric properties of health related quality of life measures in acute coronary syndrome patients: a systematic review protocol. JBI Database System Rev Implement Rep. 2018; 16, 316-23. doi: 10.11124/JBISRIR-2016-003044

26. Lima TM, Aguiar PM, Storpirtis S. Evaluation of quality indicator instruments for pharmaceutical care services: A systematic review and psychometric properties analysis. Res Social Adm Pharm. 2018; 14, 405-12. doi: 10.1016/j.sapharm.2017.05.011

27. Moher D, Liberati A, Tetzlaff J, Altmann DG.The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6(7): e1000097. doi: 10.1371/journal.pmed1000097

28. Terwee CB, Jansma EP, Riphagen II, De VeT H. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009; 18,1115-1123. doi: 10.1007/s11136-009-9528-5

29. Hanel P, Vione C. Do student samples provide an accurate estimate of the general public? PLoS One. 2016; 11(12): e0168354. doi:10.1371/journal.pone.0168354

30. Peterson RA, Merunka DR. Convenience samples of college students and research reproducibility. Journal of Business Research. 2014; 67, 1035-41. doi: 10.1016/j.jbusres.2013.08.010

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

31. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016; 5, 210.doi: 10.1186/s13643-016-0384-4

32. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 7th ed. Edinburgh Gate, Harlow: Pearson Education Limited. 2014.

33. Pedrosa I, Suárez-Álvarez J, García-Cueto E. Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. Acción Psicológica. 2013; 10, 3-18. doi: 10.5944/ap.10.2.11820

34. Terwee CB, Bot SDM, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2014; 60(1), 34-42. doi: 10.1016/j.jclinepi.2006.03.012

35. Barton KL, Wrieden WL, Anderson AS. Validity and reliability of a short questionnaire for assessing the impact of cooking skills interventions. J Hum Nutr Diet. 2011; 24, 588-595. doi: 10.1111 / j.1365-277X.2011.01180.x

36. Bech-Larsen T, Tsalis G. Impact of cooking competence on satisfaction with food-related life: Construction and validation of cumulative experience & knowledge scales. Food Quality and Preference. 2018; 68,191-197. doi: 10.1016/j.foodqual.2018.02.006

37. Begley A, Paynter E, Dhaliwal SS. Evaluation tool development for food literacy programs. Nutrients. 2018; 10(11), 1617. doi: 10.3390/nu10111617

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

39. Lavelle F, McGowan L, Hollywood L et al. The development and validation of measures to assess cooking skills and food skills. Int J Behav Nutr Phys Act. 2017; 14(1), 118. doi: 10.1186/s12966-017-0575-y

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

40. Martins CA, Baraldi LG, Scagliusi FB, Villar BS, Monteiro, CA. Cooking Skills Index: Development and reliability assessment. Rev Nutr. 2019; Published online: 14 February 2019. doi: 10.1590/1678-9865201932e180124

41. Michaud P. Development and evaluation of instruments to measure the effectiveness of a culinary and Nutrition education program. Thesis. Clemson: Clemson University, SC. 2007.

42. Poelman MP, Dijkstra SC, Sponselee H et al. Towards the measurement of food literacy with respect to healthy eating: The development and validation of the self perceived food literacy scale among an adult sample in the Netherlands. Int J Behav Nutr Phys Act. 2018; 15(1), 1-12. doi: 10.1186/s12966-018-0687-z

43. Bell R, Marshall DW. The construct of food involvement in behavioral research: scale development and validation. Appetite. 2003; 40, 235-244. doi:10.1016/s0195-6663(03)00009-6

44. Beaton DE, Bombardier C, Guillemin F, Ferraz, MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000; 25(24), 3186-3191. doi: 10.1097/00007632-200012150-00014

45. Sobral A, Freitas C, Pedroso M et al. Definições Básicas: Dado, Indicador e Índice In: Saúde Ambiental: Guia Básico para a Construção de Indicadores, pp.25-52 [Freitas CMd, editor] Brasília, DF: Ministério da Saúde. 2011. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/saude_ambiental_guia_basico.pdf10.1590/1980-549720190041

46. TenHouten WD. Scale gradient analysis: A statistical method for constructing and evaluating Guttman Scales. Sociometry. 1969; 32, 80-98.

47. DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. Qual Quant. 2018; 52, 1523-59. doi: 10.1007/s11135-017-0533-4

48. World Health Organization. Guidelines on submitting research proposals for ethics review. 2018. Available from: https://www.who.int/ethics/review-committee/guidelines/en/ (accessed December, 2019)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

49. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects p. 373-4. 2001.

50. Guttman L. A basis for scaling qualitative data. American Sociological Review. 1944; 9, 139-50. doi:10.2307/2086306

51. Menzel H. A new coefficient for scalogram analysis. Public Opin Q. 1953; 17, 268–280.

52. Jobling D, Snell EJ. The use of the coefficient of reproducibility in attitude scaling. The Incorporated Statistician. 1961; 11, 110-8.

53. Polit DF. Getting serious about test–retest reliability: a critique of retest research and some recommendations. Qual Life Res. 2014; 23, 1713–1720. doi: 10.1007/s11136-014-0632-9

54. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008; 65(23), 2276-84.

55. McDonald M. Systematic Assessment of Learning Outcomes: Developing Multiple-Choice Exams. Boston: Jones and Bartlett Publishers. 2001.

56. Sireci SG. The construct of content validity. Soc Indic Res. 1998; 45, 83-117. doi: 10.1023/A:1006985528729

57. Cronbach, LJ, & Meehl, PE. Construct validity in psychological tests. Psychol Bull. 1955; 52, 281–302. doi:10.1037/h0040957

58. Hagger MS, Gucciardi DF, Chatzisarantis NLD. On nomological validity and auxiliary assumptions: The importance of simultaneously testing effects in social cognitive theories applied to health behavior and some guidelines. Front Psychol. 2017; 8, Article 1933. DOI: 10.3389/fpsyg.2017.01933

59. Damasio B. Uso da análise fatorial exploratória em psicologia. Avaliação Psicológica. 2012; 11, 213-28.

60. Matsunaga M. How to factor-analyze your data right: do’s, don’ts, and how-to’s. International Journal of Psychological Research. 2010; 3, 97-110. doi:10.21500/20112084.854

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

61. Santos RO, Gorgulho BM, Castro MA, Fisberg R, Marchioni DM, Baltar VT. Análise de Componentes Principais e Análise Fatorial: diferenças e similaridades na aplicação em Epidemiologia Nutricional. Rev Bras Epidemiol. 2019; 22, e190041. doi: 10.1590/1980-549720190041

62. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. Psicologia Reflexão e Crítica. 2017; 30, 3. doi: 10.1186/s41155-016-0057-1

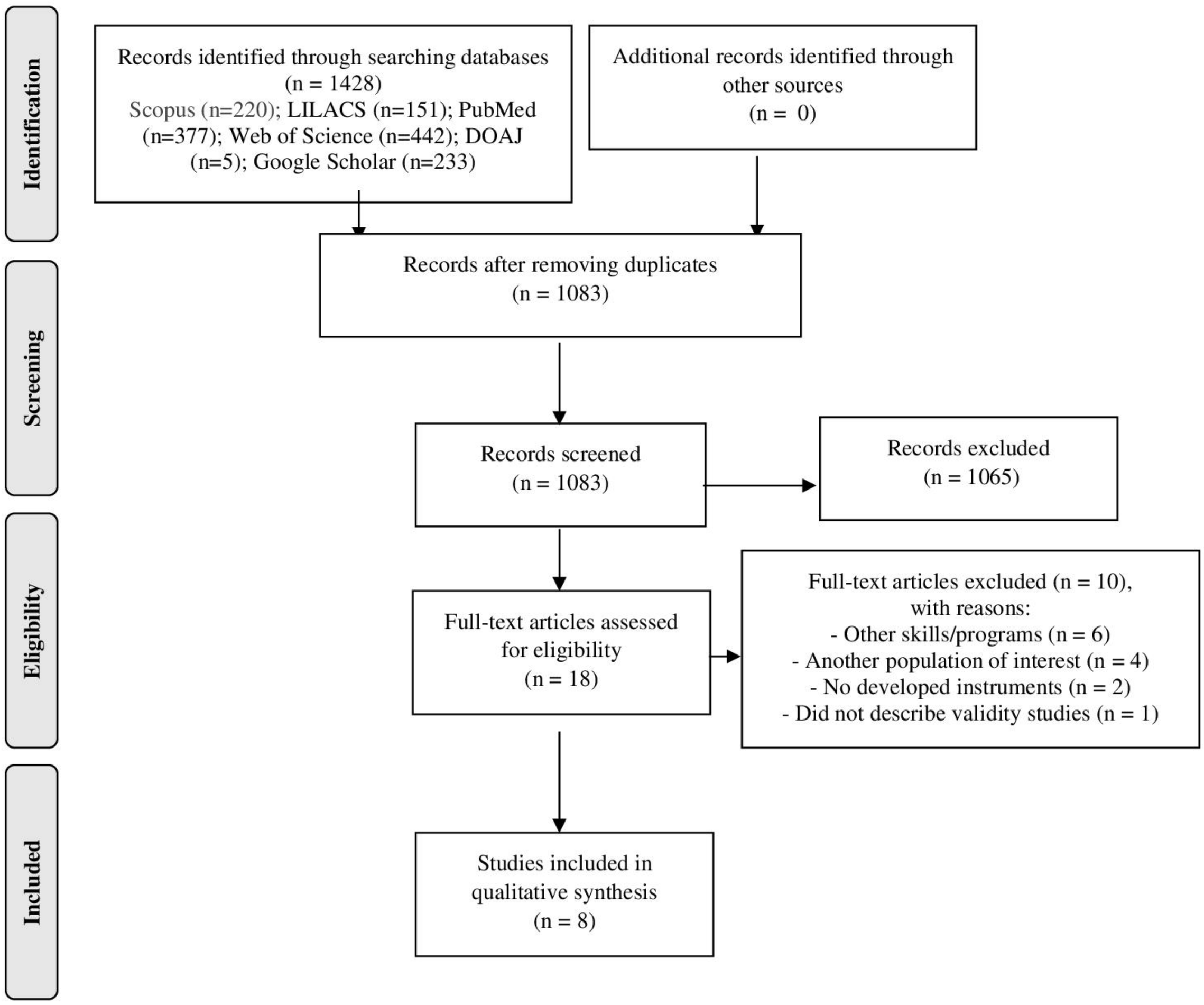
Supporting Information

S1 Appendix. Systematic review protocol. From: Aline Rissatto Teixeira, Daniela Bicalho, Tacio de Mendonça Lima. Evidence for the validation quality of culinary skills instruments: a systematic review. PROSPERO 2019 CRD42019130836. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019130836

S1 Table. PRISMA 2009 Checklist. From: Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

S2 Appendix. Search strategy until June 13, 2019.

S2 Table. List of excluded studies.



medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted June 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .