medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 in adults: What is the quality 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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 August 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 are important objects of study in the field of Public Health. Studies that propose to develop instruments for assessing such construct show lack of methodological uniformity to report validity and reliability of their instruments.

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

Design: We conducted a systematic review according to the PRISMA statement. We searched literature PubMed/Medline, Scopus, LILACS, and Web of Science databases until January 2021, and consulted Google Scholar for relevant grey literature. Two reviewers independently selected the studies, conducted data extraction, and assessed the psychometric quality of the instruments. A third reviewer resolved any doubts or disagreements in all steps of the systematic review.

Results: The search identified 1148 potentially relevant studies, out of which 9 met the inclusion criteria. In addition, we included 3 studies by searching the related articles and the reference lists of these studies, totaling 12 included studies in this review. Ten studies reported the development of tools measuring culinary skills in adults and 2 studies performed cross-cultural adaptations of original instruments. We considered adequate quality of internal consistency reliability in four studies. One study received adequate rating for test-retest reliability. No studies presented adequate rating for content validity and four studies showed satisfactory results for at least one type of construct validity. One study reported criterion validity and the quality of this psychometric property was inadequate.

Conclusions: We identified many studies that surveyed culinary skills. Although the isolated measures appraised in this review show good promise in terms of quality of psychometric properties, no studies presented adequate measures for each aspect of reliability and validity. A more consistent and consensual definition of culinary skills 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 culinary skills.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 [2-6], however, there is no consensus on its definition or a consistent theoretical debate about it [3]. This systematic review considers a broad definition of culinary skills proposed by De Oliveira, 2018 [7], as a set of attributes related to the selection and combination of foods and the use of culinary procedures and utensils involved in the planning, organization and preparation of “from scratch” meals based on fresh, minimally processed foods and culinary ingredients.

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 relates 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 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 [7].

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

an emancipatory competence associated with a healthy food routine [10]. 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 harmful effects on consumers' health, such as overweight, obesity, cancer and other chronic diseases and addiction-like behavior [11; 12]. It is worth mentioning that culinary practices also relate 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 [13].

The main source of cooking knowledge and skills is through parents [14; 15; 16; 17]. This information highlights the importance of adult cooking skills as a role model in food preparation habits development in children. In addition, Sidenvall *et al.* (2001) [18] 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 [14; 17; 19]. Among the publications on this subject are studies that propose to develop instruments that measure culinary skills in adults through the analysis of their psychometric properties.

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 [20-22]. 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 [21]. Validity refers to the fact that a tool measures exactly what it proposes to measure. It is based on extent theory research and experts' judgement (content validity), the degree in which a group of variables represents the construct to be

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

measured (construct validity) and the degree in which the instrument is related to some external criterion, considered a widely accepted measure (criterion validity) [21-23].

There are public health policies focused on cooking in several parts of the world [3]. Despite the importance of developing instruments that measure culinary skills as a strategy to assist the planning food and nutrition educational actions based on culinary practices, studies have shown lack of methodological uniformity to report validity and reliability of their instruments.

McGowan *et al* (2014) [24] conducted a review of the literature relating to the composition and measurement of an individual's domestic Cooking Skills (CS) and Food Skills(FS), providing a conceptual and critical analysis of existing measures, and reported on associations of CS and FS with dietary outcomes. However, searches were limited to journal articles in English and limited psychometric data was available in the included studies. Furthermore, the subject of food practices in public health is rapidly evolving, and other culinary skills measurement tools are likely to have been published since they reviewed the literature in 2014.

Additionally, previous reviews have not proposed to appraise the quality of psychometric properties of instruments measuring culinary skills, 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. This 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 in adult population, and critically appraise the quality of their psychometric properties.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

Methods

We registered the protocol of this systematic review on the International Prospective Register of Systematic Reviews (PROSPERO database; http://www.crd.york.ac.uk/PROSPERO/; registration number CRD42019130836). The protocol is available in the S1 Appendix. The PRISMA [25] (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

We performed a comprehensive literature search for articles published until January 12, 2021, 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*. [26]. The S2 Appendix shows the full search strategy for all databases. In addition, we conducted a grey literature search in Google Scholar to identify studies not indexed in the databases listed above. We also evaluated references to the articles found, in order to include any potential studies not yet identified.

Study selection

This review, included articles meeting the following criteria: 1) address culinary skills in adults; 2) describe the instrument's validation and reliability process, which can be original or adapted instruments. No filters for year of publication, country or language were employed. Articles that developed original instruments or reporting cross cultural adaptation of instruments addressed to measure culinary skills in children and adolescents or those whose instruments were not available (in the article or upon request to the authors) were excluded. For initial screening of abstracts and titles, we used the *Rayyan Web Platform for Systematic Reviews* [27]. Two authors (A.R.T. and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

D.B.) independently screened the titles and abstracts of citations to identify potentially relevant studies. We obtained and reviewed the full-text articles for further assessment according to the inclusion and exclusion criteria. When we could not obtain the full text, we contacted the corresponding authors by e-mail or other tools, such as ResearchGate (www.researchgate.net). A third reviewer (T.M.L.) resolved any doubts or disagreements between the reviewers regarding the inclusion or exclusion of articles. The third reviewer compared the results of the independent selection of articles carried out by the two reviewers. If the third reviewer identified any differences, he would ask the two authors to discuss their opinions. If the two reviewers did not reach an agreement, the third reviewer would present his opinion.

Data extraction and analysis

Two authors (A.R.T. and D.B.) independently performed data extraction using a preformatted spreadsheet in Microsoft Excel. A third reviewer (T.M.L.) resolved any disagreements or doubts resolved any disagreements or doubts occurred in this step, by comparing the data extraction carried out by the two reviewers. If the third reviewer identified any differences, he would ask the two authors to discuss their interpretations. If the two reviewers did not reach an agreement, the third reviewer would present his opinion. We also consulted the third reviewer in case of any doubts regarding the inclusion of potentially relevant articles identified during this step of the systematic review.

The information extracted consisted of descriptive data of the study (country, phenomenon studied, participants, sample size, instrument format, target public, number of items and domains of the instrument, development methodology and statistics performed to report psychometric properties).

Quality of psychometric properties

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

We determined the psychometric quality according to the rating system adapted from Hair Jr, Black, Babin *et al*.[28]; Pedrosa, Suárez-Álvarez, and García-Cueto [29]; and Terwee *et al.* (2007) [30]. The criteria addressed the following properties: a) reliability, including internal consistency and stability; b) validity, including content, construct (structural, hypothesis testing and cross-cultural) and criteria. We reported the quality of each measurement property as adequate (+), indeterminate (?), inadequate (-), or no information available (0). When the appraisal of the quality of a specific attribute was not applicable, we reported as 'NA'. Table 1 shows the quality criteria for psychometric properties. Two independent authors (A.R.T. and D.B.) applied this rating system, and the third reviewer (T.M.L.) resolved any divergences between them (i.e. no consensus on the rating regarding the quality of measure of an instrument), by comparing the results of critical appraisal of the quality of psychometric properties of the instruments, carried out by the two reviewers. If the third reviewer identified any differences, he would ask the two authors to discuss their opinions. If the two reviewers did not reach an agreement, the third reviewer would present his opinion.

Table 1. Quality criteria for psychometric properties of measurement (adapted from Hair Jr *et al.* [28], Pedrosa *et al.* [29], and Terwee *et al.* [30]).

Property	Definition	Ranking**	Quality criteria
Content Validity	Extent to which the domain of interest is comprehensively sampled by the items in the instrument	+	Clear description provided about the aims of the instrument, the target population, the concepts being measured, the item selection AND clear description about the experts involved in the items selection with adequate number of professionals integrating experts panel (>3) [a]; *AND* Quantitative approach for content validity: use of statistical methods derived from the experts' judgment: o Content Validity Index (CVI) = good CVI value >0.8[b]; o Critical Content Validity Ratio (CVR critical) based on the number of experts required to agree an item essential [c] (established at 0.05 significance)
		?	Lack of clear description about the aspects mentioned above; *OR* Only target population involved; *OR* Lack of information regarding the number of professionals integrating experts panel; *OR* Incomplete evaluation (no quantitative approach for content validity).
		-	No target population involved; *OR* Quantitative evaluation: CVI value < 0.8 [b] ; *OR* CVRc (critical CVR) not based on the number of experts required to agree an item essential [c] OR inappropriate level of significance to ensure unlikely random agreement for CVRc.
		0	No information found on content validity
Construct Validity	Extent to which a set of measured variables actually represents the theoretical latent construct those variables are designed to measure	+	Structural Validity: o Exploratory factor analysis OR Principal Component Analysis (PCA) performed on adequate sample size (minimum ratio of 5:1 AND/OR > 100) AND Bartlett's sphericity test (p < 0.005) OR KMO adequacy test (≥0.7) AND factors explaining ≥60% of the variance AND high factor loads, indicating that they converge to a common point, according to sample size[d]; *AND/OR* o Confirmatory factor analysis with RMSEA ≤0.07 OR GFI and AGFI ≥0.95 OR SRMR ≤0,08 OR CFI ≥0.95[d]; *AND/OR* Hypothesis testing (convergent/discriminant validity): *OR* o Convergent validity: Presents a suitable comparator instrument, providing clear information about what the comparator instrument(s) measure(s) AND shows high correlation (>0.7) between the means scores with the comparator instrument measuring related constructs[e]; *OR* o Discriminant validity: Presents a suitable comparator instrument, providing clear information about the unrelated construct the comparator instrument(s) measure(s); AND shows weak correlation (<0.3) between the means scores with the comparator instrument measuring unrelated constructs[e]; *OR* o Discriminant validity by Known (extreme) groups: performed with clear description of two or more groups expected to have different levels of the construct AND T-test performed for independent samples, reporting means differences between groups considering p-value <0.05[b]; *AND/OR* Cross Cultural Validity (when applicable): o Clear description of cross-cultural stages[f]: 1) Translations into the target language by minimum of two translators (one informed and one uninformed); 2) Synthesis of the Translations; 3) Back translation into the original language by minimum of two translators with source language as their mother tongue, unaware and uninformed of the concepts explored; 4) Submission to experts committee, with reported consensus for semantic, idiomatic, experiential and conceptual equivalences between the source and target version; 5) Pretest (minimum of 30-40 participants, with evidence if samples were similar for relevant characteristics); 6) Submission of documentation to the developers or coordinating committee for appraisal of the adaptation process.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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.

		?	Structural Validity: o No exploratory factor analysis performed OR no Principal Component Analysis (PCA) performed AND/OR no Confirmatory Factor Analysis performed; OR unclear description of methods and sample size mentioned above; *AND/OR* Hypothesis testing (convergent/ discriminant validity): o Convergent/ Discriminant validity: Lack of clear description about the information mentioned above for convergent or discriminant validity; OR insufficient information regarding the correlation results with the comparator instrument; *OR* o Discriminant validity by Known (extreme) groups: Lack of clear description of groups expected to have different levels of the construct; *OR* insufficient information regarding the T-test results; *AND/OR* Cross Cultural Validity (when applicable): o Lack of clear description about all the stages mentioned above.
		-	Structural Validity: o Factor analysis (exploratory or confirmatory) (*OR* Principal Component 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 *OR* low factor loads according to sample size[d] *OR* o Confirmatory factor analysis with RMSEA >0.07 *OR* GFI and AGFI <0.95 *OR* SRMR >0,08 *OR* CFI <0.95[d] *AND/OR* Hypothesis testing (convergent/discriminant validity): o Convergent validity: Correlation with an instrument measuring related construct <0.70[e]; *OR* o Discriminant Validity: correlation with an instrument measuring unrelated construct >0.3[e]; *OR* o Discriminant validity by Known (extreme) groups: T-test performed for independent Known (extreme) groups samples, reporting means differences between groups considering p-value >0.05; *AND/OR* Cross Cultural Validity (when applicable): o Translation and/or back translation led by 1 translator
		0	No information found on Structural Validity; *AND/OR* No information found on Hypothesis testing (convergent/discriminant validity); *AND/OR* Cross Cultural Validity (when applicable): Not informed.
Criterion Validity	Extent to which the scores of an instrument relate to the scores of a gold standard measurement	**+**	Convincing arguments that gold standard is ''gold'' *AND* correlation between change scores calculated with results ≥0.70[b;h].
		?	No convincing arguments that gold standard is ''gold'' OR unclear design or method.
		-	Correlation with gold standard <0.70, despite adequate design and method[b; h].
		0	No information found on criterion validity.
Internal Consistency reliability	Extent to which items in a scale are intercorrelated	**+**	Cronbach's alpha between 0.70 and 0.95 with mean correlation values between the items > 0.30[b, h]; *OR* Composite Reliability (CR)>0.7[i];

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

	(consistency among the variables)	**?**	Unclear design or method (e.g. reporting adequacy with lack of clear description about the statistics mentioned above).
		-	Cronbach's alpha <0.70 or >0.95 or mean correlation values between the items <0.30, despite adequate design and method [b, h]; *OR* Composite Reliability (CR)< 0,7 [i].
		0	No information on internal consistency reliability.
Stability (test-retest reliability)	Extent to which the instrument is stable over time, given by the agreement among individuals who are evaluated twice	+	Evidence provided that test conditions were similar; *AND* Intraclass correlation coefficient (ICC) *OR* Kappa ≥0.70[g]; *AND* Adequate interval (10 to 14 days)[i] between test and retest and at least 50 subjects to be considered adequate sample[b,g].
		?	Unclear design or method (e.g., time interval not mentioned *OR* inadequate time interval *OR* no evidence regarding test conditions, *OR* inadequate sample).
		-	Intraclass correlation coefficient (ICC) *OR* Kappa ≤0.70[g], despite adequate design, method and interval.
		0	No information found on stability.

** +: adequate ?: indeterminate adequacy; -: inadequate; 0: no information available. NA: Not Applicable

Abbreviations: ICC: Intraclass Correlation Coefficient; CVI: Content Validity Index; CVR critical: Critical Content Validity Ratio; 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.

[a] Pedrosa *et al.*,2013 [29]; [b] Souza, Alexandre, Guirardello , 2017[21]; [c] Ayre & Scally, 2014 [31]; [d] Hair Jr. *et al.*, 2014 [28]; [e] Abma, et al, 2016[32]; [f] Beaton *et al.,* 2000 [33]; [g] Polit, 2011 [34]; [h] Terwee *et al.*, 2007 [30]; [i] Valentini, & Damasio, 2016 [35]; [j] Keszei *et al.*, 2010 [36].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 August 9, 2021. 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 .

Results

Search results

The electronic search (including gray literature databases) identified 1148 potentially relevant studies. After reviewing the titles and abstracts, we selected 16 articles for full-text examination. Of these, nine studies (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al*., 2011 [39]; Jomori *et al*, 2017 [40]; Barton *et al*, 2011 [41]; Kowalkowska *et al*, 2018 [42]; Lavelle *et al*, 2017 [43]; Kennedy *et al*, 2019 [44]; Martins *et al*, 2019[45]) met the inclusion criteria for review. A list of the excluded studies is available in the S2 Table. The authors presented only one doubt during the selection and data extraction processes, which was resolved by the third reviewer. The doubt corresponded to the inclusion of a potentially relevant article identified during the full text reading of the articles (Hartmann *et al*, 2013[46]). We identified other two relevant studies, by searching the reference lists of the included studies (Vrhovnik, 2012 [47]; Condrasky *et al*, 2013 [48]). Finally, 12 studies were included in this systematic review. Fig. 1 shows a flowchart of the literature search.

Fig 1. Study selection flowchart of literature search. Abbreviations LILACS: Latin American and Caribbean Health Sciences Literature.

Characteristics of the studies

Studies were carried out in the United States of America (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al*, 2011 [39]; Condrasky *et al*, 2013 [48]), Brazil (Jomori *et al*, 2017 [40]; Martins *et al*, 2019 [45]), Canada (Vrhovnik, 2012 [47]; Kennedy *et al*, 2019 [44]), Switzerland (Hartmann *et al*, 2013 [46]), Portugal (Kowalkowska *et al*, 2018 [42]), Scotland (Barton *et al*, 2011 [41]) and Northern Ireland and Republic of Ireland (Lavelle *et al*, 2017 [43]). All of them were

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

published in English, between 2007 and 2019. One study did not seek ethical approval (Barton *et al*, 2011 [41]).

Included papers had distinct purposes: those reporting the development of an original instrument, or cross-cultural adaptation of a tool to explicitly measure cooking/food skills or a part thereof (n = 7) and original tools developed to evaluate a cooking and food skills intervention (n = 5). Most tools assessed cooking skills in adults from a particular country (Hartmann *et al*, 2013 [46]; Lavelle *et al*, 2017 [43]; Vrhovnik, 2012 [47]; Kennedy *et al* l, 2019 [44]), parents of schoolchildren responsible for food preparation at home (Martins *et al*, 2019 [45]), university students (Warmin, 2009 [38]; Jomori *et al*, 2017 [40]; Kowalkowska *et al*, 2018 [42]) and adults of low-income communities, participants in culinary skills and nutrition education programs (Michaud, 2007 [37]; Condrasky *et al*, 2011 [39]; Condrasky *et al*, 2013 [48]; Barton *et al*, 2011 [41]). Study samples were mostly composed of women (Barton *et al*, 2011 [41], Condrasky *et al*, 2011 [39]; Martins *et al*, 2019 [45]; Kennedy *et al*, 2019 [44]; Kowalkowska *et al*, 2018 [42]; Michaud, 2007 [37]; Warmin, 2009 [38]; Vrhovnik, 2012 [47]). The participants' age ranged from 18 to 69 years. Sample sizes ranged from 29 to 4.4306 individuals.

Table 2 shows the characteristics of the included studies.

239 **Table 2. Characteristics of the included studies.**

Author/year/ Country	Objective	Instrument (mnemonic)	Latent phenomenon evaluated	Study sample	Age years/ range /mean(SD)	N
Michaud / 2007 /USA	To provide evidence for and demonstrate the processes used to develop and test tools to measure the effectiveness of a culinary and nutrition education program	Cooking and healthy eating evaluation questionnaire	Cooking skills and healthy eating, based on the main objectives of the Cooking with a Chef (CWC) program	Experts panel: professionals (nutrition, public health, gastronomy, sociology, statistic); Pilot data and larger study: Parents and caregivers recruited from preschool, public school, church, and organized playgroup settings in South Carolina.	Pilot and larger study: 18 to 50 years old.	Experts panel (n=12); Pilot data (n = 39), with test-retest subgroup (n = 19; Larger study data (n = 162).
Warmin / 2009 / USA, article published in 2012	To test the effects of an established culinary nutrition program with college students and to test the effectiveness of placing of the nutrition component onto an online presentation.	Cooking and healthy eating evaluation questionnaire (online version; new items)	Cooking skills and healthy eating, based on the main objectives of the Cooking with a Chef (CWC) program	College students recruited from a Nutrition for non-majors class offered through Clemson University's Food Science and Human Nutrition Department, who received no 'Cooking with a Chef' intervention.	18 to 22 years old.	Test-retest (n=29)
Condrasky et al. / 2011 / USA	To develop scales to assess the impact of the Cooking with a Chef program on several psychosocial constructs	Cooking and healthy eating evaluation questionnaire (short version)	Cooking skills and healthy eating, based on the main objectives of the Cooking with a Chef (CWC) program	Experts panel: Academic professionals; Pilot and larger study: Parents, caregivers and cooks, largely female recruited from child care settings in South Carolina and church and school kitchens.	Pilot and larger study: 35 years old or older.	Experts panel (n=4); Pilot data (n = 39), with test-retest subgroup (n = 19); Larger study data from self-selected parents and caregivers (n = 162) and cooks (n=83);
Condrasky et al. / 2013 / USA	To develop and evaluate a participatory training for cooks in African American churches.	Cooking and healthy eating evaluation questionnaire (new items)	Cooking skills and healthy eating, based on the main objectives of The Faith, Activity, and Nutrition (FAN) program , adapted from the CWC program	Experts panel: Nutrition professionals; Pilot data: cooks and planning committee members.	Not specified.	Not specified.
Jomori et al. /2017 / Brazil	To describe the results of the construct validity by the known-groups'method of a Brazilian cooking skills and healthy eating questionnaire	Brazilian version of the Cooking and healthy eating evaluation questionnaire	Cooking skills and healthy eating	Students who had started their undergraduate program at Federal University of Santa Catarina (UFSC), Brazil in 2015 were selected based on convenience. They voluntarily accessed the online questionnaire from August to November 2015.	20.7 (±5.59) years old.	University students (n=767).
Hartmann et	To design a	Cooking	Cooking skills	Adult participants from the Swiss Food Panel, a	55.5 years (SD = 14.6,	test-retest: n= 4436.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

al / 2013 / Switzerland	cooking skill scale that is reliable and applicable to most people	Skills Scale (CSS)		population-based longitudinal study of the eating behavior of the Swiss population.	range 21–99).	
Kowalkowska et al. /2018 /Portugal	To assess the reliability of a Portuguese version of the Cooking Skills Scale (CSS) and to evaluate the association between cooking skills and socio-demographic, psychological and other cooking related variables	Portuguese version of the Cooking Skills Scale (CSS)	Cooking skills	Portuguese university students, attending bachelor's or integrated master's degree studies, with access to cooking facilities.	22.8 years (SD= 4.9; range: 17 - 58).	Larger study data (n=730); Repeatability - test-retest (n= 106).
Barton et al. / 2011 / Scotland	To undertake an assessment of validity and reliability of a short questionnaire designed to measure the impact of cooking skills interventions on cooking confidence, the use of basic food skills, and food selections amongst low income communities	Short Questionnaire - CookWell programme	Cooking confidence and food skills, based on the key domains shown to be influenced by the CookWell programme	Experts panel: dietitians and community workers; Face validity: adults residing in Tayside, Scotland, who were typical of those who may attend cooking skills classes; Reliability: group of adults attending community-based classes (other than cooking) in Tayside, Scotland; Feasibility: from participants from the 'Get Cooking' project, living in areas in the lower deciles (most deprived) of the Scottish Index of Multiple Deprivation.	Face validity: range of 21-69 years; Reliability: 46 (15.1); Feasibility: 35.0 (20.8).	Experts panel (n= 28); Face validity (n=20); Reliability (n=57); Feasibility (n =24).
Vrhovnik / 2012/ Canada	To create a valid and reliable tool to assess the level of food skills in the community	Food skills survey tool	Food skills	Face validity: public health dieticians and nurses; Content validity: Field experts in food skills and survey development from Queen's University, the research department at KFL&A (Kingston, Frontenac, Lennox & Addington Health Unit); Pilot, factor validity and reliability: adults from Kingston, Frontenac, Lennox and Addington counties, able to understand English, recruited through the directories of residential phone numbers provided by CCI Research.	Pilot, factor validity and reliability: age>18 years old.	Face validity (n=13); Content validity (n=10); Pilot, factor validity and reliability (n=-273).
Lavelle et al. / 2017 / Ireland	To develop and validate a measure for cooking skills and one for food skills, that are clearly described, relatable, user-friendly,	Cooking skills confidence measure, and Food skills	Cooking skills and food skills	Experts' opinion: professionals working in the area of health promotion including cooking and food skills interventions and education; Study 1: adults responsible for preparing a main	Study 1: 20–60 years old; Study 2: 18 -27 years old;	Study 1 - content and convergent validity (n=1049); Study 2 - test-retest

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

	suitable for different types of studies, and applicable across all sociodemographic levels.			meal at least once per week; Study 2: students from Ulster University enrolled on a course that consisted of nutrition, hospitality, food marketing or food product and innovation orientated modules; Study 3: students from the Ulster University, Northern Ireland and St. Angela's College Sligo, Ireland, either studying a Business-related degree or were studying Home Economics, classified as 'Food preparation novices' and 'Experienced food preparers'; Study 4: combination of the samples in study 2 and study 3 (representing the P/P method) and participants randomly selected from the sample in Study 1 ((representing the CAPI method).	Study 3: 'Experienced food preparers': 18–26 years old; 'Food preparation novices':19–24 years old.	and internal consistency reliability of the measures in the P/P format (n = 23); Study 3 - discriminant validity and further assess the internal consistency reliability of the P/P measures(n = 57); Study 4 - differences between the CAPI and the P/P method in relation to the confidence scores of the measure (n= Studies 2 and 3 + 38 from study 1).
Kennedy et al / 2019/ Canada	To develop, validate, and assess reliability of a food skills questionnaire	Food skills questionnaire	Basic to intermediate food skills	Content validity: Dietitians, home economists, academics, and chefs; Face validity: convenience sample of students at Western University; Test-retest and inter-item reliability: undergraduate students randomly selected.	Test-retest and item reliability: (mean age, 22 ± 6 years.	Content validity (n=17); Face validity (n=20) Test-retest and inter-item reliability (n=165; lowest number of participants answering the same questions at times 1 and 2 was 126).
Martins et al / 2019/ Brazil	To describe the development and the reliability assessment of an index that evaluates the confidence in performing cooking skills considered relevant in Brazil.	Cooking Skills Index (CSI)	Cooking skills self-efficacy	Face validity and experts panel: nutritionists, physicians and biologists belonging to the research group that supported the preparation of the Dietary Guidelines for the Brazilian Population; Pilot, test- retest and internal consistency: parents of schoolchildren responsible for food preparation at home.	Not specified.	Face validity and experts panel (n= 6); Pilot (n=10) and test-retest and internal consistency (n=51).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 August 9, 2021. 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 .

Characteristics of the instruments

All studies provided description of the construct, with conceptual framework or clear rationale to define their instruments' construct.

Six studies reported the development (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al*, 2011 [39]; Condrasky *et al*, 2013 [48]; Barton *et al*, 2011 [41]) or cross-cultural adaptation of tools (Jomori *et al.*, 2017 [40]) aiming to evaluate cooking skills and healthy eating, based on the main objectives of cooking and nutrition education interventions programs.

Michaud (2007) [37], developed an original questionnaire, consisting of 51 items measuring culinary skills and healthy eating, aiming to evaluate the Cooking with Chef (CWC) intervention Program. In 2009, Warmin [39] tested the online format of application of this questionnaire, based on a sample of university students. In addition, Condrasky *et al.* (2011) [39] reported the alteration of three scales in this questionnaire. Condrasky *et al.* (2013) [48] then adapted a few items of the questionnaire, employed in a sample of church cooks in South Carolina (USA). The final questionnaire consisted of 64 items, with one knowledge evaluation section, a short index and six scales related to Self-Efficacy for produce consumption, cooking, using basic techniques, using fruits, vegetables, and seasonings during cooking practices. Finally, Jomori *et al.* (2017) [40] described the results of a cross-cultural adaptation of the later version of the culinary skills and healthy eating questionnaire for Brazilian students and reported its construct validity.

Barton *et al.* (2011) [41] also described the results of the development and validation of a short cooking skills questionnaire, aiming to evaluate the effects of the CookWell intervention program. The questionnaire consisted of 19 items with domains related to frequency of preparing food, confidence in following a simple recipe, cooking with basic ingredients, and preparing new foods and recipes. Some items were similar to those reported in the cooking skills and healthy eating questionnaire, based on the main objectives of the Cooking with a Chef (CWC) program, described in the aforementioned studies (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al.*, 2011 [39]; Condrasky *et al.*, 2013 [48]; Jomori *et al.*, 2017 [40]).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 remaining studies (Hartmann *et al.*, 2013 [46]; Kowalkowska *et al.*, 2018 [42]; Lavelle *et al.*, 2017 [43]; Vrhovnik, 2012 [47]; Kennedy *et al.*, 2019 [44]; Martins, *et al.*al, 2019 [45]) described the results of the development and validation or cross-cultural adaptation of tools aiming to evaluate adults' cooking and/or food skills or a part thereof, with some similarities. The instruments' domains ranged from 1 to 3 and the number of items ranged from 7 to 39, mainly related to 'food preparation techniques', 'meal planning' and 'food selection and purchase'. We present the main characteristics of these instruments, their domains and items in common as well as the divergences below.

The Cooking Skills Scale (originally developed by Hartmann *et al.*, 2013 [46]; and adapted for Portuguese university students, by Kowalkowska *et al.*, 2018 [42]) focused on measuring cooking skills, based on the ability to prepare certain dishes and products, but without distinguishing whether they were prepared with basic ingredients, pre-prepared products, convenience foods or a combination of them.

Lavelle *et al.* (2017) [43] developed measurements to assess cooking skills and food skills confidence. The cooking skills confidence measure consisted of 2 domains: 'Food preparation Techniques' and 'Cooking method', including items related to skills for cooking pre-prepared products and convenience foods (e.g.: *rate how good they are at: Microwave food, including heating ready-meals*).

Unlike the items shown in Hartmann *et al.*'s cooking scale (2013) [46] and Lavelle *et al.*'s cooking confidence measure (2017) [43], the Cooking Skills Index, developed by Martins, *et al.*, (2019) [45] focused on cooking self-efficacy related to the preparation of meals from the combination of natural or minimally processed foods and seasoned using natural seasonings and culinary ingredients.

Lavelle *et al.*'s food skills confidence measure (2017) [43] consisted of five domains related to meal planning, shopping, budgeting, resourcefulness and label reading. Kennedy *et al.*'s food skills questionnaire (2019) [44] focused on similar domains, such as 'Food Selection and Planning',

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

'Food Safety and Storage'; however, it comprised one domain related to 'Food Preparation'. Like Martins *et al.*'s [45] instrument for accessing cooking skills (2019), this domain included items to assess confidence in performing cooking techniques, (e.g.: *rate your confidence in boiling, steaming or stewing*) and using basic ingredients and seasoning (e.g.: *rate your confidence in: preparing food from basic ingredients; choosing a spice or herb*). Vrhovnik's Food skills survey tool (2012) [47] also consisted of domains regarding 'Mechanical Techniques' (using texture, taste and smell to guide cooking methods), 'Food Preparation' (chopping, mixing, blending, cooking and following recipe) and 'Conceptualizing Foods' (creating meal ideas with leftover food and adjusting recipes to fit the needs of an individual).

The studies reported analysis of the psychometric properties of their instruments: Only two studies presented statistical methods derived from the experts' judgment for content validity (Kennedy *et al.*, 2019 [44], Vrhovnik, 2012[47]). Six out of ten studies that proposed to develop and validate a new instrument (or those reporting small changes in the original tool) did not report construct validity (Warmin, 2009 [38]; Condrasky *et al.*, 2013 [48], Hartmann *et al.*, 2013 [46]; Barton *et al.*, 2011[41]; Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]). Two studies reported cross-cultural adaptation of instruments (Kowalkowska *et al.*, 2018 [42], Jomori *et al.*, 2017 [40]). Only one study (Michaud, 2007 [37]) reported criterion validity. Nine studies tested the reliability of their instrument according to internal consistency (Michaud, 2007 [37]; Condrasky *et al.*, 2011 [39]; Jomori *et al.*, 2017 [40]; Kowalkowska *et al.*, 2018 [42]; Barton *et al.*, 2011 [41]; Vrhovnik, 2012 [47]; Lavelle *et al.*, 2017 [43], Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]) and/or stability (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al.*, 2011 [39]; Barton *et al.*, 2011 [41]; Kowalkowska *et al.*, 2018 [42]; Lavelle *et al.*, 2017 [43], Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]; Hartmann *et al.*, 2013 [46]). Table 3 describes the characteristics of the instruments.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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.

317 **Table 3. Characteristic of the instruments**

Author, year	Instrument	No. of items	Target population	Format	Domains evaluated (number/name)	Develop-ment	Psychometric Properties				
							Content Validity	Construct Validity	Criterion Validity	Internal Consisten-cy	Stability reliability
Michaud, 2007	Cooking and healthy eating evaluation questionnaire	51	Low-income adult parents and caregivers of preschool-age children in South Carolina attending the Nutrition Education Program, with Hands-On Cooking Activities, Cooking With a Chef (CWC)	Survey applied in written form, consisted of six scales, one knowledge evaluation section, and a short index developed to evaluate the Cooking with Chef (CWC) intervention: 1 Availability and Accessibility of Fruits and Vegetables Index - yes/no; 1 Knowledge evaluation (test); Scales: 5-point Likert scales	Availability and Accessibility of Fruits and Vegetables Index (8 items); Knowledge evaluation (8 tests); 6 scales: Cooking Behavior Scale (CB; 3 items); Self efficacy scales: Produce Consumption Self-Efficacy Scale (SEPC; 3 items); Cooking Self-Efficacy Scale (SEC; 6 items); Self-Efficacy for Using Basic Techniques Scale (SECT;12 items); Self-Efficacy for Using Fruits, Vegetables, and Seasonings Scale (SEFVS; 4 items); Cooking Attitude Scale (CA; 7 items)	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: No QA: No	SV: Yes HT: No CCV: NA	GS: No PV:Yes CcV: No	yes	yes
Warmin, 2009	Cooking and healthy eating evaluation questionnaire (online version; new items)	62	College-aged students (majority between the ages of 18 and 20) attending the Nutrition Education Program, with Hands-On Cooking Activities, Cooking With a Chef (CWC) - with face-to-face and online intervention	Online survey consisted of six scales, one knowledge evaluation section, and a short index: adapted from the cooking skills and health eating evaluation questionnaire developed to evaluate the Cooking with Chef (CWC) intervention (Michaud, 2007).	Availability and Accessibility of Fruits and Vegetables Index (8 items); Knowledge evaluation (8 tests); 6 scales: Cooking Behavior Scale (CB; 10 items); Self efficacy scales : Produce Consumption Self-Efficacy Scale (SEPC; 3 items); Cooking Self-Efficacy Scale (SEC; 6 items); Self-Efficacy for Using Basic Techniques Scale (SECT;12 items); Self-Efficacy for Using Fruits, Vegetables, and Seasonings Scale (SEFVS; 8 items); Cooking Attitude Scale (CA; 7 items)	DC: Yes CF: Yes TP: Yes DSS: No	EP: No TP: No QA: No	SV: No HT: No CCV: NA	GS: No PV: No CcV: No	No	Yes
Condrasky et al., 2011	Cooking and healthy eating evaluation questionnaire (short version)	22	Low-income adult parents/ caregivers and cooks in South Carolina attending the Nutrition Education Program, with Hands-On Cooking Activities, Cooking With a Chef (CWC)	Three 5- point Likert scales of the cooking skills and health eating evaluation questionnaire developed to evaluate the Cooking with Chef (CWC) intervention program, primarily validated by Michaud (2007) were changed.	3 scales: Negative Cooking Attitude (4 items); Self-efficacy for eating/ cooking fruit and green or root vegetables (4 items); Self-efficacy for cooking techniques and meal preparation (14 items)	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: Yes QA: No	SV: Yes HT: No CCV: NA	GS: No PV: No CcV: No	Yes	Yes
Condrasky et	Cooking and healthy eating	64	Church kitchen staff members	Online survey consisted of 5-point scales, one knowledge	Knowledge evaluation (8 tests);	DC: Yes CF: Yes	EP: Yes TP: No	SV: No HT: No	GS: No PV: No	No	No

al., 2013	evaluation questionnaire (new items)		attending the cook's training of the Faith, Activity, and Nutrition (FAN) program (adapted from the CWC program) in South Carolina	evaluation section, and a short index, adapted from the cooking skills and health eating evaluation questionnaire proposed by Warmin et al., 2009 and originally developed by Michaud, 2007. 2-item Pre/Post Training Evaluation added (measured on a 10-point scale), not validated in previous studies.	Availability and Accessibility of Fruits and Vegetables Index (8 items); 6 scales: Cooking Behavior Scale (CB; 10 items); Self efficacy scales : Produce Consumption Self-Efficacy Scale (SEPC; 3 items); Cooking Self-Efficacy Scale (SEC; 6 items); Self-Efficacy for Using Basic Techniques Scale (SECT;12 items); Self-Efficacy for Using Fruits, Vegetables, and Seasonings Scale (SEFVS; 8 items); Cooking Attitude Scale (CA; 7 items) Pre/Post Training Evaluation related to cooking skills and confidence in meal preparation: (2 items)	TP: Yes DSS: No	QA: No	CCV: NA	CcV: No		
Jomori et al., 2017	Brazilian version of the Cooking and healthy eating evaluation questionnaire	51	Brazilian university students	Cross-cultural adaptation of the cooking skills and health eating evaluation questionnaire developed to evaluate the Cooking with Chef (CWC) intervention (Michaud, 2007).	Availability and Accessibility of Fruits and Vegetables Index (8 items); Knowledge evaluation (8 tests); 6 scales: Cooking Behavior Scale (CB; 3 items); Self efficacy scales : Produce Consumption Self-Efficacy Scale (SEPC; 3 items); Cooking Self-Efficacy Scale (SEC; 6 items); Self-Efficacy for Using Basic Techniques Scale (SECT;12 items); Self-Efficacy for Using Fruits, Vegetables, and Seasonings Scale (SEFVS; 4 items); Cooking Attitude Scale (CA; 7 items)	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: NA TP: NA QA: NA	SV: No HT: Yes CCV: Yes	GS: No PV: No CcV: No	Yes	No
Hartmann et al., 2013	Cooking Skills Scale (CSS)	7	Swiss adult population	Paper-and-pencil questionnaire; Six-point scale, from strongly disagree (1 point) to strongly agree (6 points)	1 domain: Cooking skills (7 items)	DC: No CF: Yes TP: Yes DSS: Yes	EP: No TP: No QA: No	SV: No HT: No CCV: NA	GS: No PV: No CcV: No	No	Yes
Kowalkowska et al., 2018	Portuguese version of the Cooking Skills Scale (CSS)	7	Portuguese university students	Cross-cultural adaptation of the Cooking Skills Scale developed by Hartmann et al. (2013); Self-administered questionnaire developed on the Lyme Survey platform; Six-point scale, from strongly disagree (1 point) to strongly agree (6 points)	1 domain: Cooking skills (7 items)	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: NA TP: NA QA: NA	SV: Yes HT: No CCV: Yes	GS: No PV: No CcV: No	Yes	Yes
Barton et al.,2011	Short Questionnaire - CookWell programme	19	Adult members of low-income communities participants in the	Two-page questionnaire utilizing a closed-question format designed based on key domains known to be influenced by	5 domains: Confidence in using a recipe; Frequency of using basic ingredients for preparation of meals;	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: Yes QA: No	SV: No HT: No CCV: NA	GS: No PV: No CcV: No	Yes	Yes

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

			CookWell programme.	cooking skills programs.	Buying less convenience food; Increased likelihood of tasting and experimenting with new foods; Fruit and vegetable consumption.						
Vrhovnik, 2012	Food skills survey tool	31 questions (18 ordinal items selected for factor analysis)	Community members/ generic population - created for use by public health units aiming to assess food skills in their respective communities	Survey delivered via telephone; Set number of Likert-scale type questions ranging from strongly disagreeing to strongly agreeing	3 domains: Mechanical techniques; Food Preparation; Conceptualizing foods.	DC: Yes CF: Yes TP: No DSS: Yes	EP: Yes TP: Yes QA: Yes	SV: Yes HT: No CCV: NA	GS: No PV: No CcV: No	Yes	No
Lavelle et al., 2017	Cooking skills confidence measure, and Food skills confidence measure	Cooking skills confidence measure: 14 Food skills confidence measure: 19	Adult population of all sociodemographic levels	Two presentation methods: Computer Assisted Personal Interviewing (CAPI) and the traditional paper and pen (P/P) styled format; The confidence score is the sum of the 1 to 7 ratings for the skills stated as used. If a skill is not used, it is scored a zero for that skill.	2 measures: Cooking skills confidence; Food skills confidence.	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: Yes QA: No	SV: Yes HT: Yes CCV: NA	GS: No PV: No CcV: No	Yes	Yes
Kennedy et al., 2019	Food skills questionnaire	39	Young adults	Online survey	3 domains: Food Selection and Planning (9 items); Food Preparation (20 items); Food Safety and Storage (10 items).	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: Yes QA: Yes	SV: No HT: No CCV: NA	GS: No PV: No CcV: No	Yes	Yes
Martins et al., 2019	Cooking Skills Index (CSI)	10	Adults responsible for food preparation at home, in São Paulo, Brazil	Telephone survey conducted by trained interviewers. Four-point scale: (0) not confident, (1) little confident, (2) confident, and (3) very confident.	1 domain: Cooking skills self-efficacy	DC: Yes CF: Yes TP: Yes DSS: Yes	EP: Yes TP: Yes QA: No	SV: No HT: No CCV: NA	GS: No PV: No CcV: No	Yes	Yes

Abbreviations:

- Development column: DC: Clear description of the construct – Yes/No; CF: Theory or conceptual framework or clear rationale provided to define the construct – Yes/No; TP: Clear description of the target population for which the instrument was developed – Yes/No; DSS: Development study sample representing the target population – Yes/No
- Content validity column: EP: Use of Experts panel for item selection – Yes/No; TP: target population involved – Yes/No; QA: Quantitative approach for content validity – Yes/No
- Construct validity column: SV: Structural Validity – Yes/no; HT: Hypothesis testing (convergent validity) – Yes/No; CCV: Cross Cultural Validity – Yes/No/ NA (Not Applicable
- Criterion validity column: GS: Use of gold standard measure for comparison – Yes/No; PV: Predictive Validity – Yes/not applicable; CcV: Concurrent Validity - – Yes/not applicable)
- Internal consistency column - Yes/No; Reliability column - Yes/No

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 August 9, 2021. 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

We describe the quality of the psychometric properties of the instruments in Table 4.

We considered adequate quality of internal consistency reliability in four studies (Condrasky *et al.*, 2011 [39]; Kowalkowska *et al.*, 2018 [42]; Lavelle *et al.*, 2017 [43]; Martins *et al.*, 2019 [45]) Three studies (Michaud, 2007 [37]; Jomori *et al.*, 2017 [40]; Kennedy *et al.*, 2019 [44]) showed inadequate quality of this measure. Two studies had the internal consistency reliability considered indeterminate: Barton *et al.*, 2011 [41] did not test three out of five sections of their instrument for internal consistency reliability; Vrhovnik, 2012 [47] did not report any statistical results, despite the author's affirmation on satisfactory results for internal consistency. Three studies (Warmin, 2009 [38]; Condrasky *et al.*, 2013 [48]; and Hartmann *et al.*, 2013 [46]) did not report internal consistency reliability.

Nine studies reported test-retest reliability (Michaud, 2007 [37]; Warmin, 2009 [38] Condrasky *et al.*, 2011 [39]; Hartmann *et al.*, 2013 [46]; Kowalkowska *et al.*, 2018 [42]; Barton *et al.*, 2011 [41]; Lavelle *et al.*, 2017 [43], Kennedy *et al.*, 2019 [44]; and Martins *et al.*, 2019 [45]). However, we considered the quality of this measure inadequate in two of these studies (Kowalkowska *et al.*, 2018 [42]; Martins *et al.*, 2019 [45]), since they presented results inferior to the minimum criterion for *Kappa,* despite adequate design and method. Five studies showed inadequate time interval (Barton *et al.*, 2011 [41]; Hartmann *et al.*, 2013 [46]) or inadequate sample size for test-retest reliability analysis (Michaud, 2007[37]; Warmin, 2009 [38]; Condrasky *et al.*, 2011 [39]; Lavelle et. al, 2017 [43]); therefore, we deemed the quality of stability inadequate in these studies.

No studies reporting the development or small changes of an original instrument provided adequate measures to show content validity. The authors did not calculate any index of agreement for content validity (Martins *et al.*, 2019 [45]; Lavelle *et al.*, 2017 [43]; Barton *et al.*, 2011 [41], Condrasky *et al.*, 2011 [39]; Condrasky *et al.*, 2013 [48]; Michaud, 2007 [37]), or statistical results of expert's agreement did not reach minimum criteria to be considered valid (Vrhovnik, 2012 [47];

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

Kennedy *et al.*, 2019 [44]). Moreover, one study did not perform any analysis of content validity (Hartmann *et al.*, 2013 [46]).

Regarding construct validity, six studies reported at least one kind of analysis (structural validity, hypothesis testing or cross-cultural validity, when applicable).

Five studies performed structural validity analysis (Michaud, 2007 [37]; Condrasky *et al.*, 2011 [39]; Vrhovnik, 2012 [47]; Kowalkowska *et al.*, 2018 [42]; Lavelle *et al.*, 2017 [43]). We classified two of them as inadequate according to the quality of this attribute, due to insufficient sample size for Exploratory Factor Analysis (EFA) (Michaud, 2007 [37]) and retention of items showing cross loadings (Lavelle *et al.*, 2017 [43]). In addition, one study performed exploratory factor analysis, however did not provide results for factor loadings (Vrhovnik, 2012 [47]), hence, we considered indeterminate quality of structural validity. Only two studies reported hypothesis testing for construct validity. One of them properly performed convergent validity with satisfactory results (Lavelle *et al.*, 2017 [43]). Jomori *et al.*, 2017 [40] performed discriminant validity between known groups, however we rated the quality of this attribute inadequate, since the authors reported no significant differences between groups in one scale (*Self-Efficacy for Using Basic Cooking Techniques (SECT)*). Two studies performed cross-cultural adaptations of original instruments (Jomori *et al.*, 2017 [40] and Kowalkowska *et al.*, 2018 [42]). One of them received inadequate rating due to insufficient number of translators leading back translation, and stages for cross-cultural adaptation were incomplete (Kowalkowska *et al.*, 2018 [42]). Six studies did not report any kind of analysis to evidence construct validity (Warmin, 2009 [39]; Condrasky *et al.*, 2013 [48]; Hartmann *et al.*, 2013 [46]; Barton *et al.*, 2011 [41]; Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]).

Most studies did not provide information on criterion validity. Only one study (Michaud, 2007 [37]) performed analysis for criterion validity, however, the authors did not describe it clearly (convincing arguments for gold standard).

Table 4. Evaluation of quality of psychometric properties of the instruments.

Instrument	Reliability		Validity				
				Construct			
	Internal consistency	Stability	Content	Structural	Hypothesis Test	Cross Cultural	Criterion
Cooking and healthy eating evaluation questionnaire (Michaud, 2007)	(-) Cronbach alpha <0,7 for CB and SEPC	(?) Inadequate sample size	(?) No quantitative approach for content validity	(-) EFA: No KMO adequacy results and inadequate sample size	(0)	NA	(?) Predictive validity: no convincing arguments that gold standard is ‘‘gold’’
Cooking and healthy eating evaluation questionnaire - online version; new items (Warmin, 2009)	(0)	(?) Inadequate sample size	(0)	(0)	(0)	NA	(0)
Cooking and healthy eating evaluation questionnaire - short version (Condrasky et al., 2011)	(+)	(?) Inadequate sample size	(?) No quantitative approach for content validity	(?) PCA: No KMO adequacy results	(0)	NA	(0)
Cooking and healthy eating evaluation questionnaire - new items (Condrasky et al., 2013)	(0)	(0)	(?) No quantitative approach for content validity	(0)	(0)	NA	(0)
Brazilian version of the Cooking and healthy eating evaluation questionnaire (Jomori et al.,2017)	(-) Cronbach alpha <0,7 for CB and CA	(0)	NA	(0)	(-) Known groups’ method: No differences (p-value > 0.05) between the means of men and women’s cooking skills in the SECT* scale	(+)	(0)
Cooking Skills Scale (CSS) (Hartmann et al., 2013)	(0)	(?) Inadequate time interval (1 year difference)	(0)	(0)	(0)	NA	(0)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

Portuguese version of the Cooking Skills Scale (CSS) (Kowalkowska et al., 2018)	(+)	(-) Cohen's Kappa (0.49)	NA	(+) PCA	(0)	(-) Back translation led by 1 translator; no submission to experts committee; Documentation to the developers for appraisal of the adaptation process was not reported	(0)
Short Questionnaire - CookWell programme (Barton, et al., 2011)	(?) Two out of five sections tested	(?) Inadequate time interval (1 week)	(?) No quantitative approach for content validity	(0)	(0)	NA	(0)
Food skills survey tool (Vrhovnik, 2012)	(?) Authors tested 1 section of the instrument, which they considered reliable. No statistical results on Cronbach's alpha were shown.	(0)	(-) Items with CVI <0,8 were reworded, however, CVI was not retested after changes were made.	(?) EFA. No KMO adequacy results and factor loadings were not reported.	(0)	NA	(0)
Cooking skills confidence measure, and Food skills confidence measure (Lavelle, et al., 2017)	(+)	(?) Inadequate sample size; Only conducted for P/P format: 23 participants	(?) No quantitative approach for content validity	(-) Items with cross loadings were retained	(+)	NA	(0)
Food skills questionnaire (Kennedy, et al., 2019)	(-) Chronbach alpha < 0,7 for the 'food safety and storage' domain	(+)	(-) Four items retained in the Food Preparation domain showing low Lawshe's CVR	(0)	(0)	NA	(0)
Cooking Skills Index (CSI) (Martins, et al., 2019)	(+)	(-) Weighted Kappa <0,7	(?) No quantitative approach for content validity	(0)	(0)	NA	(0)

*Abbreviations: SECT: Self-Efficacy for Using Basic Cooking Techniques

Rating: (+) = adequate; (?) = indeterminate; (-) = inadequate; (0) = no information available; NA= Not applicable

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 August 9, 2021. 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 .

Discussion

Summary of evidence

To our knowledge, this is the first systematic review to identify and appraise quality of psychometric properties of instruments for assessing culinary skills in adults. This article has provided a comprehensive critical analysis of the studies' characteristics and their psychometric properties. We found twelve studies developing original instruments to measure culinary skills in adults, or performing cross-cultural adaptations.

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 of culinary skills measurements. Although we considered adequate quality of internal consistency reliability in four studies, only one study received adequate rating for stability (test-retest reliability). No studies developing original instruments presented satisfactory measurement for content validity since the authors did not calculate any index of agreement. Only four studies showed satisfactory results for at least one type of construct validity (structural, hypothesis testing or cross- cultural adaptation, when applicable) and only one study reported criterion validity, however, we considered inadequate quality of this measurement property. These results indicate that although there are isolated measures appraised in this review that show good promise in terms of quality of psychometric properties, no studies presented satisfactory results for each aspects of reliability and validity.

General view of the studies

Most studies are originally from countries whose native language is English. One Brazilian study (Martins *et al.*, 2019 [45]) originally developed an instrument in Portuguese for application with parents of schoolchildren, responsible for food preparation at home. However, the authors translated the instrument from Brazilian Portuguese into English, without making it available in its

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

original language. Despite the authors' intention to provide access to their study using universal language, translating the instrument into 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 [33].

Regarding submission of psychometric studies for ethical approval, one study (Barton *et al.*, 2011 [41]) justified the absence of ethical approval because it comprised developmental work for service evaluation. Despite the fact that validation studies aim at the development of tools for measuring latent phenomena, methods applied to report 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 [49; 50].

General view of the instruments

Most studies reported the development of scales, indexes, and questionnaires. One study classified their instrument as an index (Martins *et al.*, 2019 [45]); however, the instrument used Likert scale to register 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 [51].

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

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 [20].

The majority of the included studies presented instruments with items assessing cooking self-efficacy (regarding food preparation techniques), meal planning and food selection and purchase. The main difference between the instruments referred to the conceptualization of culinary skills: some authors comprehend that such skills comprise the ability to prepare certain dishes, including those based on pre-prepared products and convenience foods (Hartmann *et al.*, 2013 [46]; Kowalkowska *et al.*, 2018 [42]; Lavelle *et al.*, 2019 [43]). However, relying on pre-prepared or convenience products to prepare a meal may require less cooking abilities [10]. Thus, using the microwave oven for the mere heating of frozen meals, for example, could overestimate an individual's skill level. Moreover, pre-prepared products and convenience foods are often classified as ultra-processed foods, whose negative nutritional attributes are associated with harmful effects to health [52]. 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.

Some authors identify cooking skills as a distinct construct from food skills. Lavelle *et al.* (2017) [43] define cooking skills as a set of physical or mechanical skills used in the production of a meal while food skills are described as a wider set of skills involved in the entirety of the meal preparation process that includes meal planning, shopping, budgeting, resourcefulness, and label reading. Short [6] however, specifies that reducing cooking skills to the ability to do tasks such as baking, broiling, poaching, and stir-frying is an oversimplification of activities involved in planning, organizing, and preparing a meal. She also states that our confidence in cooking and using basic skills influences what and how we cook, which may influence our diet quality. Kennedy *et al.* (2019) [44] seem to consider mechanical skills for meal production as part of the overall construct of food skills. The authors also state that low food skills or cooking self-efficacy are barriers to

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

healthy eating. Vrhovnik (2012) [47] conceptualizes food skills as necessary abilities for knowledge, planning, conceptualization, preparation and perception of food. Although these authors quoted such domains to define the construct of food skills, they seem to be aligned with the concept of cooking skills adopted in this review [7], reinforced by Short [6].

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 reliability (Michaud, 2007 [37]; Condrasky *et al.*., 2011 [39]; Jomori *et al.*, 2017 [40], Kowalkowska *et al.*, 2018 [42], Barton *et al.*, 2011 [41]; Lavelle *et al.*, 2017 [43]; Vrhovnik, 2012 [47], Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]). Internal consistency reliability 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 [28].

Three studies showed insufficient results for Cronbach's alpha (Kennedy *et al.*, 2019 [44]; Michaud, 2007 [37]; Jomori *et al.*, 2017 [40]). Two of them were studies aiming to validate the cooking skills CWC questionnaire: Michaud's evaluation tool (2007) [37] showed inadequate Cronbach's alpha coefficients for the *Cooking Behavior scale (CB scale)*. Similar results were observed in Jomori *et al.*'s (2017) [40] study: The *Cooking Behavior (CB*) and *Cooking Attitude (CA)* scales showed low internal consistency reliability. The later authors argued that problems in the process of cross-cultural adaptations concerning translation of the original instrument into Brazilian Portuguese might have occurred. The items corresponding to these scales might not represent the constructs the authors intended to measure [28]. Thus, it is important to adjust these items for more appropriate translation into Brazilian Portuguese and to perform factor analysis [30].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

Barton *et al*. (2011) [41] did not test three out of five sections of their questionnaire, under the justification that the domains within each section of the instrument assessed different constructs. Vrhovnik (2012) [47] tested only one section of her instrument to report internal consistency reliability, which the author affirmed to be reliable; however, no statistical results were shown. Therefore, we deemed these studies inadequate, according to internal consistency reliability quality criteria presented in this review.

We considered indeterminate quality of stability reported in six studies, due to insufficient sample size or inadequate time interval to perform test- retest reliability (Michaud, 2007 [37]; Warmin, 2009 [38]; Condrasky *et al.*, 2011 [39]; Barton *et al.*, 2011 [41]; Hartmann *et al.*, 2013 [46]; Lavelle *et al.*, 2017 [43]). Hartmann *et al.* (2013) [46] performed the test-retest within 1-year time interval, which may result in a measurement error to show the instrument's stability and reproducibility [23]. Participants might improve their culinary skills during the interval between the test and the retest, especially if the elapsed time is too long. We also observed insufficient *Kappa* values (<0.7) in two studies that reported test-retest reliability. Therefore, we rated inadequate quality of this attribute.

Studies that relied exclusively on internal consistency reliability and stability analysis, without performing other psychometric measurements to validate their 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) [34;53]. Six studies fit into this scenario (Warmin, 2009 [38]; Condrasky *et al.*, 2013 [48]; Hartmann *et al.*, 2013 [46]; Barton *et al.*, 2011 [41]; Kennedy *et al.*, 2019 [44]; Martins *et al.*, 2019 [45]). The authors only reported results for Cronbach's alpha and test-retest reliability, and conducted inappropriate analysis for content validity, disregarding empirical evidence for experts' agreement (or did not perform any tests for content validity). Moreover, these studies did not present any tests for construct or criterion validity.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

All studies aiming to develop and validate an original instrument failed to show proper content validity: most studies relied on face validity, literature research, and experts' judgment; however, the authors did not calculate any index to confirm experts' agreement. Content validity based on 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 [22; 28; 39, 54].

We evaluated the quality of construct validity measures of studies reporting structural validity (Michaud, 2007 [37]; Condrasky *et al.*, 2011 [39]; Vrhovnik, 2012 [47]; Lavelle, *et al.*, 2017 [43]; Kowalkowska, *et al.*, 2018 [42]), hypothesis testing (Lavelle *et al.*, 2017 [43]; Jomori *et al.*, 2017 [40]) or cross-cultural validity for adapted instruments (Jomori *et al.*, 2017 [40]; Kowalkowska *et al.*, 2018 [42]). We observed a number of limitations, according to the quality criteria for this attribute, presented in this review.

Regarding structural validity, two studies performed principal component analysis and three studies performed exploratory factor analysis. No studies performed Confirmatory Factor Analysis (CFA). According to Gruijters, 2019 [55], exploratory factor analysis (EFA) and principal component analysis (PCA) explain correlations between items to some extent, but component analysis does a poorer job at it because it includes a portion of irrelevant variance in the analysis. If researchers have a clear idea about what a scale is supposed to be measuring, it is highly recommended studies perform Confirmatory factor analysis (CFA) to test *a priori* ideas about the latent variables researchers intend to measure [31; 30].

Only two out of five studies reporting structural validity (Lavelle *et al.*, 2017 [43]; Kowalkowska *et al.*, 2018 [42]) described the Kaiser-Meyer-Olkin (KMO) adequacy test with values > 0.80, which is considered very good for factor analysis appropriateness [28].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

Michaud (2007) [37] performed exploratory factor analysis on insufficient sample size (minimum ratio of 5:1). Costello & Osbourne, 2005 [56] caution researchers to remember that *EFA is a "large-sample" procedure and that generalizable or replicable results are unlikely if the sample is too small*.

The cross-cultural adaptation of Michaud's (2007) [37] instrument, reported by Jomori *et al.* (2017) [40] was adequately performed and showed satisfactory results. However, we considered inadequate quality of measure for discriminant validity between known groups, since the study showed unsupported results for significant differences between the means scores of one scale (*Self-Efficacy for Using Basic Cooking Techniques (SECT)*). This type of validity evaluates the presence of differences in the measurements obtained between the groups, not whether the measure actually measures the intended construct [57], hence, we suggest performing structural analysis to confirm construct validity of this instrument.

Vrhovnik (2012) did not provide statistical results for items factor loadings, which may imply inadequate decisions regarding retention or exclusion of an item [28].

Despite satisfactory results for convergent validity in Lavelle *et al.*'s study (2017) [43], the exploratory factor analysis (EFA) performed to validate the construct of the Cooking skills and the Food skills confidence measures showed some limitations. Four 'Food skill' items had higher loadings in the 'Cooking Skills' domain (*Buying in Season; Using leftovers to create another meal; Keeping Basics in the cupboard and Reading the best before date*), however they were retained in the 'Food Skills' factor. When a variable is found to have more than one significant loading, it is hard to make those factors be distinct and represent separate concepts [28]. If an instrument shows items with several cross-loadings, the items may be poorly written or the *a priori* factor structure could be flawed [58]. Moreover, two 'Cooking Skills' items fit into a third factor; however, they were left in the 'Cooking Skills' measure. One of these items consisted in '*Microwave food (not drinks/liquid) including heating ready-meals*'. The Brazilian Food Guide (2014) states that although

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

microwaving may be used in meal preparation (for example microwaving a ready meal) it is not seen as a cooking skill.

Kowalkowska *et al.*, 2018 [42] performed a cross cultural adaptation of Hartmann *et al.*'s cooking skills instrument (2013) [46]. However back translation was inadequately performed. Beaton (2000) [33] recommends minimum of two back-translators with the source language as their mother tongue. The main reasons are to avoid information bias and to increase the likelihood of highlighting the imperfections in the translated questionnaire.

Regarding criterion validity, little information was available in the included studies. Only one study (Michaud, 2007) presented criterion validity (predictive validity). However, we considered inadequate quality of this attribute. 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 [21; 59].

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. 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

This review identified many studies surveying culinary skills; we considered most instruments insufficient, according to the quality of their psychometric properties. Thus, the flaws observed in these studies show that there is a need for ongoing research in the area of the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

psychometric properties of instruments assessing culinary skills. Moreover, our findings contribute to supporting the selection of valid and reliable instruments by healthcare professionals in clinical and Public Health settings.

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 without overestimating one's skills, based on his/hers ability of heating convenience food. Considering items measuring culinary skills related to the use of using basic ingredients and seasoning proves itself essential for greater understanding of barriers and facilitators related to healthy culinary practices.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

3. 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

4. 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

5. 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.

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

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

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. 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

10. 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

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

12. Askari M, Heshmati J, Shahinfar H, Tripathi N, Daneshzad E. Ultra-processed food and the risk of overweight and obesity: a systematic review and meta-analysis of observational studies. Int J Obes (Lond). 2020 Oct;44(10):2080-2091. https://doi.org10.1038/s41366-020-00650-z.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

13. 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

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

15. 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.

16. 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.

17. 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

18. 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.

19. 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.

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

21. 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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

23. 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

24. McGowan L, Caraher M, Raats M, Lavelle F, Hollywood L, McDowell D, Spence M, McCloat A, Mooney E, Dean M. Domestic cooking and food skills: A review. Crit Rev Food Sci Nutr. 2017 Jul 24;57(11):2412-2431. doi: 10.1080/10408398.2015.1072495. PMID: 26618407.

25. 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

26. 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

27. 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

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

29. 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

30. 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

31. Ayre C, Scally AJ. Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation. Measure Eval Counsel Dev. 2014;47(1):79–86. doi: 10.1177/0748175613513808

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

32. Abma, I.L., Rovers, M. & van der Wees, P.J. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. BMC Res Notes 9, 226 (2016). doi: 10.1186/s13104-016-2034-2

33. 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

34. 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

35. Valentini F, Damásio BF. Variância Média Extraída e Confiabilidade Composta: Indicadores de Precisão. Psic.: Teor. e Pesq. 2016; 32(2):1-7. doi: 10.1590/0102-3772e322225

36. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res. 2010 Apr;68(4):319-23. doi: 10.1016/j.jpsychores.2010.01.006

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

38. Warmin, A, Cooking With a Chef: a culinary nutrition intervention for college aged students. Thesis. Clemson: Clemson University, SC. 2009.

39. Condrasky, MD; Williams, JE; Catalano, PM; Griffin, SF. Development of psychosocial scales for evaluation the impact of a Culinary Nutrition Education program on cooking and healthful eating. J Nutr Educ Behav. 2011, 43(6):511-516. doi: 10.1016/j.jneb.2010.09.013

40. Jomori, MM; Proença, RPdaC; Echevarria-Guanilo, ME; Bernardo, GL; Uggioni, PL; Fernandes, AC. Construct validity of Brazilian cooking skills and healthy eating questionnaire by the known-groups method. Br Food J. 2017, 119(5)00-00. doi: 10.1108/BFJ-10-2016-0448

41. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

42. Kowalkowska, J, Poínhos, R; Rodrigues, S. Cooking skills and socio-demographics among Portuguese university students. Br Food J. 2018, 120(3)563-577. doi: 10.1108/BFJ-06-2017-0345

43. 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

44. Kennedy LG, Kichler EJ, Seabrook JA, Matthews JI, Dworatzek PDN. Validity and Reliability of a Food Skills Questionnaire. J Nutr Educ Behav. 2019;51(7):857-864. doi: 10.1016/j.jneb.2019.02.003.

45. 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

46. 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

47. Vrhovnik, L. A pilot study for the development of a food skills survey tool. Dissertation. Queen's University. Kingston, Ontario, Canada, 2012.

48. Condrasky MD, Baruth M, Wilcox S, Carter C, Jordan JF. Cooks training for faith, activity, and nutrition project with AME churches in SC. J Eval Program Plann. 2013, 37:43-9. doi: 10.1016%2Fj.evalprogplan.2012.11.002

49. World Health Organization. Guidelines on submitting research proposals for ethics review. 2018. Available from: https://www.who.int/ethics/review-committee/guidelines/en/ (accessed January, 2021)

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

51. 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:

medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .

http://bvsms.saude.gov.br/bvs/publicacoes/saude_ambiental_guia_basico.pdf10.1590/1980-549720190041

52. Pagliai G, Dinu M, Madarena MP, Bonaccio M, Iacoviello L, Sofi F. Consumption of ultra-processed foods and health status: a systematic review and meta-analysis. Br J Nutr. 2021, 125(3):308-318. doi: 10.1017/S0007114520002688.

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

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

55. Gruijters, S. Using principal component analysis to validate psychological scales: Bad statistical habits we should have broken yesterday II. The European Health Psychologist. 2019, 20(5), 544-549. https://ehps.net/ehp/index.php/contents/article/view/3301

56. Costello, AB; Osborne, J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research, and Evaluation. 2005, 10 (7). doi: 10.7275/jyj1-4868

57. Eschevarria-Guanilo, ME; Gonçalves, NR; Juceli, P. Psychometric properties of measurement instruments: conceptual bases and evaluation methods - part II Texto Contexto Enferm. 2019, 28: e20170311. doi: 10.1590/1980-265X-TCE-2017-0311

58. Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th ed.). Boston, MA: Pearson.

59. 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

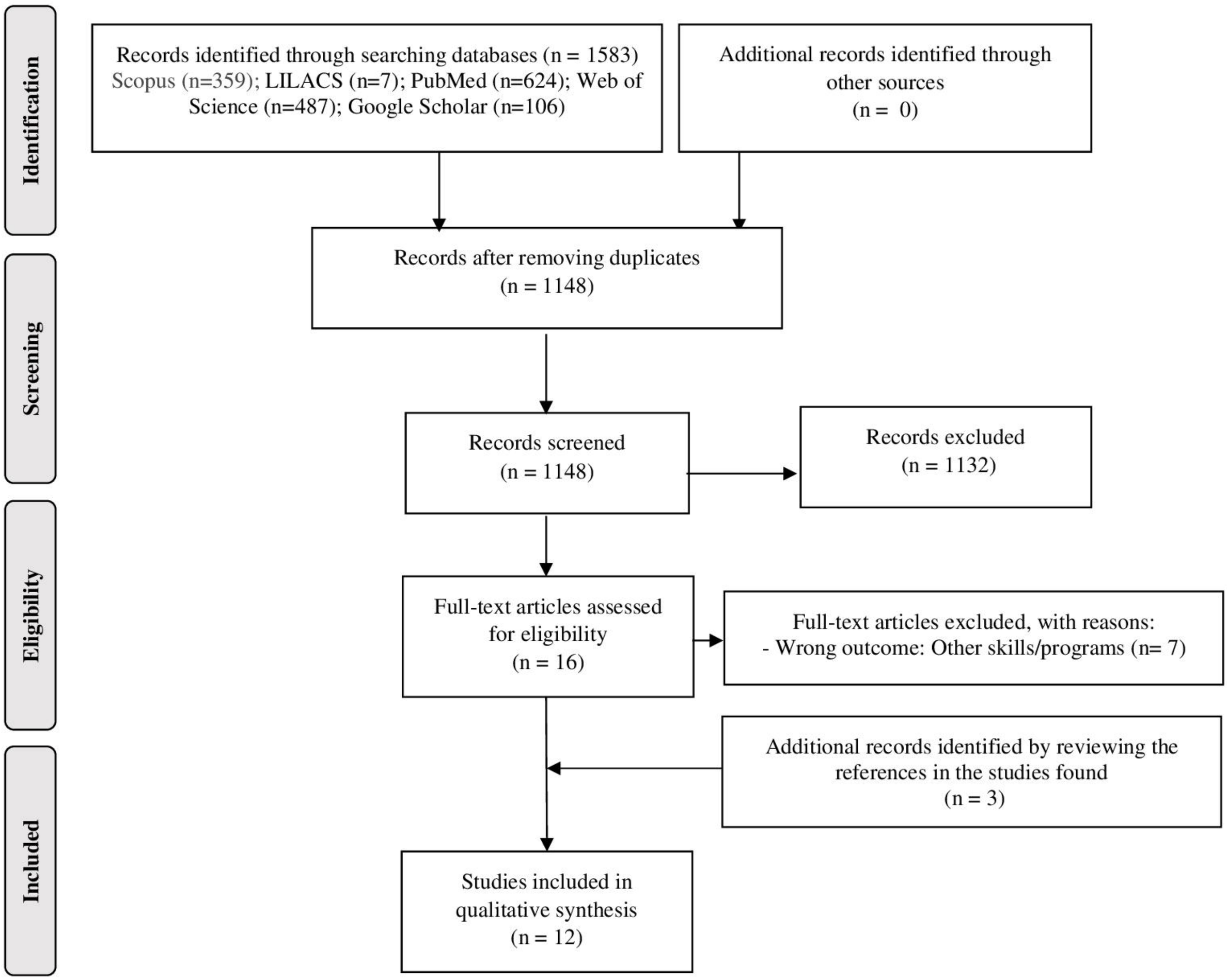
medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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 January 12, 2021.

S2 Table. List of excluded studies.



medRxiv preprint doi: https://doi.org/10.1101/2020.06.12.20129668; this version posted August 9, 2021. 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 .