Heliyon 10 (2024) e32226

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Assessing reliability and validity of food safety culture assessment tools

Shingai P. Nyarugwe, Lone Jespersen *

Faculty of Allied-Health and Wellbeing, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK

A R T I C L E I N F O

Keywords:
Food safety culture
Reliability
Validity
Methodology
Assessment tool
Trustworthiness

A B S T R A C T

Since its recognition as a plausible direction to assure food safety, food safety culture research has evolved with several commercial and scientific assessment tools developed to evaluate the food safety culture in food businesses. However, existing research does not specify the validity and reliability checks required to demonstrate rigor in the tool development process and there is no unified methodology to confirm robustness of the tools to ensure trustworthiness and usefulness of findings and inferences generated. The purpose of the study was to develop a method to evaluate food safety culture assessment tools and to assess the reliability and validity of existing food safety culture assessment tools using the developed method. Eleven elements were found to be key in validating food safety culture assessment tools. Of the eight tools assessed, only one tool (CT2) was validated on each of the elements. The depth of validation strategies differed for each tool. Three out of the five commercial tools published peer reviewed publications that demonstrated the validation checks that were done. Face validation, and pilot testing were evident and appeared to be done the most. Whilst content, ecological, and cultural validity were the least validated for scientific tools, factor analysis and reliability checks were the least evaluated for commercial tools. None of the tools were assessed for postdictive validity, concurrent validity and the correlation coefficient relating to construct validity. Having an established science-based approach is key as it provides a way to determine the trustworthiness of established assessment tools against accepted methods.

1. Introduction

Food safety culture is on the global agenda as a plausible concept to assure safety of food. Governments, third parties (e.g., standard bodies, commercial organisations), researchers, and the food industry have been focused on conceptualising and operationalising the concept. From a regulatory perspective, food safety legislation has evolved by integrating food safety culture into existing formal frameworks to ensure and verify that food businesses meet their obligation to assure food safety [1–3]. Likewise, food safety culture assessment requirements have been incorporated into third party standards [4]. Several commercial tools have also emerged e.g., VEST framework by SGS and Culture Excellence by Taylor Shannon International. Various researchers have also developed several tools to assess food safety culture in different sectors of the food industry ([5–7]; Jespersen et al., 2014, [8,9]). These different frameworks and assessments offer different levels of depth when analysing food safety culture [10].

While food safety culture research has evolved [9], with several company-specific, commercial, and scientific assessment tools

* Corresponding author
E-mail address: lone@cultivatefoodsafety.com (L. Jespersen).

https://doi.org/10.1016/j.heliyon.2024.e32226
Received 8 September 2023; Received in revised form 16 April 2024; Accepted 29 May 2024
Available online 31 May 2024

developed to evaluate the food safety culture in food businesses, current efforts do not specify a unified approach to demonstrate rigor in the tool development process. Existing approaches to food safety validation consider the validation of food safety control measures, methods of analysis, and methods for detection, identification and quantification from a food safety hazards perspective [11–13] and do not yet consider the validation of cognitive tools especially since food businesses are expected to assess their food safety culture [14]. It is therefore important to ensure the validity and reliability of the developed assessment tools and to evaluate the extent to which checks have been done to confirm their robustness. Validity and reliability are the two validation strategies widely adopted in the evaluation of measurement instruments. Validity demonstrates a system's accuracy and reliability demonstrates a system's ability to produce consistent and repeatable results [15,16]. Both reliability and validity are useful in demonstrating and communicating rigor of the research process and for demonstrating trustworthiness of the research findings [15,17]. Appropriateness of the constructs, content, scales, contexts, cultural and language constraints, relevance, understandability, how well the scale measures the intended concept, reliability, consistency, and intended respondents are some of the decisions that researchers need to make in the validation process [18].

Reliability can be demonstrated through the evaluation of internal consistency, and inter-rater, test-retest, parallel form and split half reliability [15,17]. Validity can be demonstrated through face, content, construct, concurrent, criterion, ecological, and cultural validity [19–22]. Statistical tests such as Cronbach's alpha coefficient, Pearson correlation coefficient, kappa statistics, content validation ratio, factor analysis also give a measure of validity and reliability strength [17,21]. However, it is impossible to reach 100 % validity and this can be compensated by ensuring rigor and trustworthy data through "honesty, depth, richness, extent of triangulation and objectivity of the researcher" [19].

Although validation approaches have been discussed, there is no unified framework or consistent methodology to assess the reliability and validity of food safety culture assessment tools. Jespersen, Griffiths, and Wallace [23] highlighted the importance of having a common approach to validate assessment tools and the adoption of an overarching framework to ensure quality and trustworthiness and to demonstrate transparency and rigor in the assessment process. Therefore, the objective of this research was twofold, firstly, to develop a methodology to evaluate the validity and reliability of food safety culture assessment tools; secondly, to assess the validity and reliability of existing scientific and commercial assessment tools using the developed methodology. We hypothesise that (i) there is no unified methodology for assessing reliability and validity of food safety culture assessment tools, (ii) without a standardised method for validating food safety culture assessment tools, it is difficult to assess the validity of the measurement tools and controlling the outcome, and (iii) having a science-based unified framework might be a good starting point to demonstrate rigor and ensure appropriate assessment tools are developed.

2. Methodology

2.1. Study design

A two-step study design was used to evaluate the validity and reliability of food safety culture assessment tools. The first step involved the development of a methodology to evaluate the tools and the second step involved identification of the tools and the actual assessment of the tools.

2.1.1. Developing a methodology for assessing reliability and validity of culture evaluation tools

A literature search for peer reviewed scientific articles published in English was done in Google Scholar, Web of Science, SCOPUS, Science Direct to identify existing methods used to evaluate assessment tools for validity and reliability. At least two databases have been identified to be sufficient [24]. The search was conducted between 30 August 2022 to 19 May 2023 and updated in April 2024. Articles were searched for in these databases using the following keywords: "assessment tool ", "evaluation tool ", "questionnaire", "survey", "food safety", "food safety culture", "safety, "safety culture", "culture", "organisational culture", "validity", "reliability", "validation", "development", "validation process", "validation steps", "validation methods". Generic and culture-specific assessment methods were reviewed. Methods listing the tool validation steps, describing the criteria useful to validate assessment tools and describing the relevant statistical analysis that should be included when determining the reliability and validity of an assessment tool were evaluated. Initial search was done for articles published in the last ten years and expanded to twenty based on the initial findings as proposed by Grewal, Kataria, and Dhawan [25]. Common, widely used, and relevant assessment methods useful in validating assessment tools were selected and used in developing the assessment methodology.

2.1.2. Assessing culture evaluation tools

2.1.2.1. Identifying the culture assessment tools. A list of commercial food safety culture assessments products was developed. These products were identified from a literature search and by food safety industry practitioners (n = 2) and academia (n = 4). The organisations purporting to develop these assessment products were approached via email and through the authors networks to confirm that they indeed had developed tools to assess food safety culture, to provide information on these tools and where the tools were published, to provide clarity and evidence on the validation done.

Three scientific tools for assessing food safety culture were selected for comparison based on purposive sampling. Purposive sampling was used as these tools have been validated and therefore meet the research criteria [26].

2.1.2.2. Reliability and validity assessments. Once the food safety culture assessment tools were identified and listed, publications provided by the tool developers and publications of the tools in grey, commercial, and academic literature were then reviewed and evaluated for validity and reliability using the assessment method developed in section 2.1.

2.2. Data analysis

Each of the criteria for the identified factors used to evaluate validity and reliability of the tools were assigned a score of one, and as there were 33 criteria, a maximum possible score of 33 was expected for a completely validated tool. Tiku and Pecht [27] suggested that a minimum weight factor of one can be used and an overall score can be assigned based on the evaluation and used in comparative analysis.

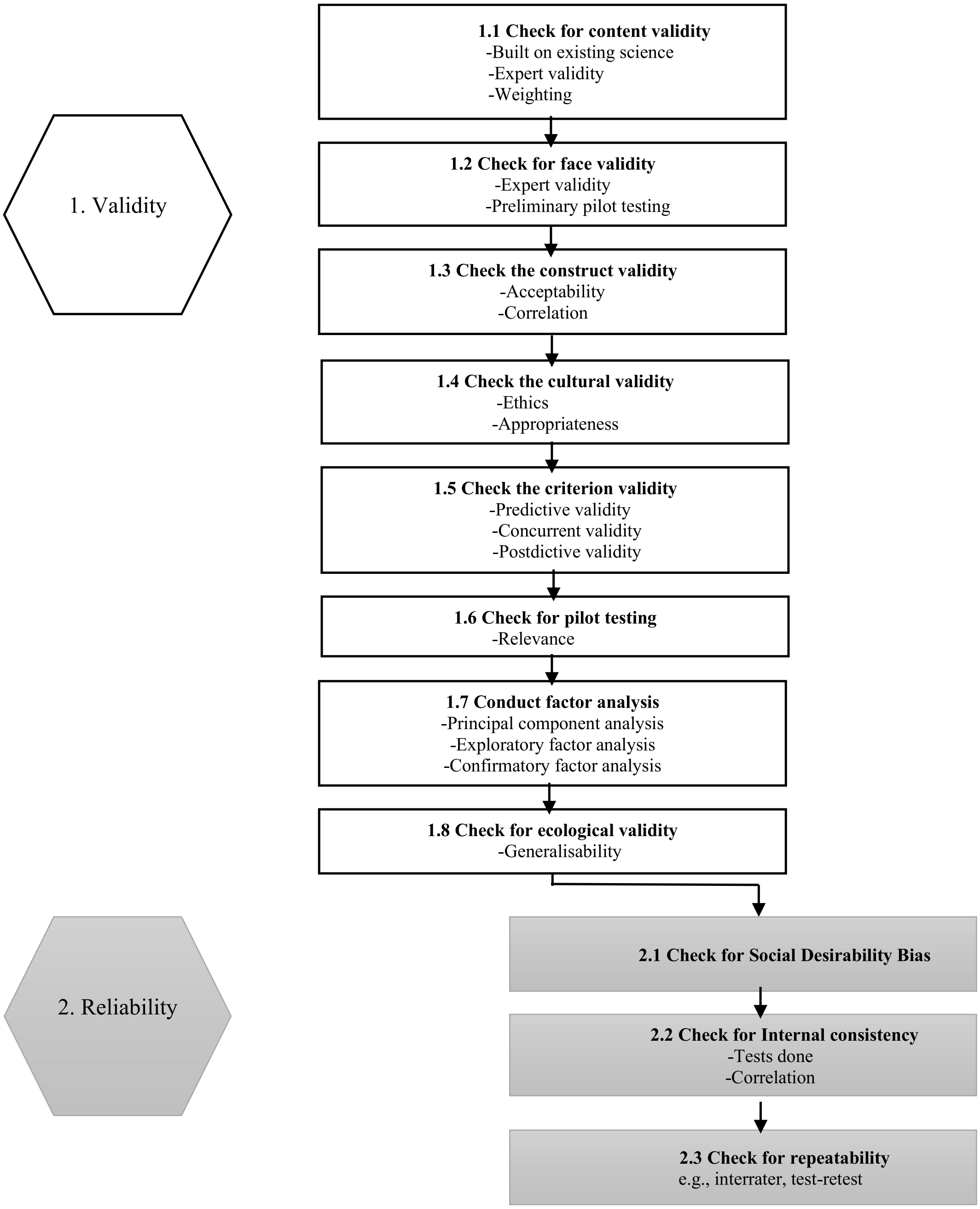

Fig. 1. Methodology used to evaluate the validity and reliability of food safety culture assessment tools.

3. Results

3.1. Validity and reliability assessment methodology

Fig. 1 shows the assessment methodology, which was developed from existing literature. Eleven factors were identified from the literature review and acknowledged to be useful to evaluate the validity and reliability of food safety culture assessment tools. Of these, eight factors were identified as key in evaluating validity and three factors to be key in evaluating reliability of the assessment tools. Each factor represents a significant element useful in validating the assessment tools. Focus areas highlight the crucial aspects to be validated. Isaksson, Eckert, Panarotto, and Malmqvist [28] argued that successful validation depends on the ability to focus on what to validate. To operationalise each factor when evaluating an assessment tool, criteria useful to characterise the factor were identified. Such criteria provide a structured way to guide the evaluation of the validity and/or reliability of an assessment tool, collect essential information about the factor being evaluated, give evidence of the actual situation and define the extent to which the assessment tools have been validated [29]. Botreau et al. [30], suggested that criteria should be exhaustive i.e., include all important aspects, minimal, i.e., contain only necessary criteria and should be independent of each other. Findings for each of the factors and relevant criteria are listed below.

3.1.1. Content validity

Content validity was identified as a key factor in validation as it ensures that the contents of the tool are evaluated to check whether they represent the theoretical construct of the tool [15]. To achieve content validity, three focus areas were identified i.e., whether the tool is built based on existing science, validated by experts and the components weighted (Fig. 1). Findings showed that whether relevant theories from other culture domains e.g., safety culture and other relevant domains such as social sciences, management and food science were considered should be evaluated. Domain definition, domain relevance, domain representation and appropriate test construction procedures were also found to be fundamental elements of test quality used in determining content validity. Additionally, the approach used for content analysis was proposed to be key in demonstrating validity.

3.1.2. Face validation

Face validation was identified to be critical in ensuring tool validation with expert validation and preliminary pilot testing identified as relevant focus areas (Fig. 1). Findings revealed that to ensure face validation, checks should be done on whether what the tool actually measures aligns with the respondent's understanding of the questions [15].

3.1.3. Construct validity

Construct validation was identified to be key in tool validation as it evaluates the extent to which an assessment tool measures what it is intended to measure [31] and involves evaluation of constructs related to the research [15]. The findings also showed that tools should demonstrate that they reliably assess multiple dimensions that reflect food safety culture [31] and that if construct validity is not done this could affect the data analysis and interpretation, and inferences made [15]. Acceptability and correlation were considered useful focus areas to assess construct validity as these measure whether the construct is an acceptable measure, and whether the scale correlates with existing tools that measure the same construct.

3.1.4. Cultural validity

Findings showed that cultural validity should contextualise validity within the community being studied [32]. Two focus areas i.e., ethics and appropriateness of the tool were found to be key to evaluate cultural validity.

3.1.5. Criterion validity

Criterion validity has also been identified as key to component of tool validation, as this demonstrates whether the tool compares with other measures or outcomes already held to be valid. The tools should therefore demonstrate "how well one measure predicts the outcome of another measure" [21,33]. Three types, namely predictive, concurrent, and postdictive validity were identified as focus areas to ensuring criterion validity.

3.1.6. Pilot testing

Pilot testing was considered important as it establishes relevance of the tool. As such, checks should be done on whether the final version of the questionnaire is tested with a considerable number of the intended respondents thus mimicking the full study [15]. Only one focus area was identified and pertains to confirming the relevance of the tool.

3.1.7. Factor analysis

Findings concluded that factor analysis is a critical component of validation as it examines "how underlying constructs influence the responses on a number of measured variables" [34]. The type of factor analysis such as exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and principal component analysis (PCA) should therefore be demonstrated. The Kaiser-Meyer-Olkin (KMO) was also considered as a crucial measure to determine how well data is suited for factor analysis.

3.1.8. Ecological validity

Evaluation of ecological validity was considered important as it aims to determine whether the tool can be generalised. Key focus areas included the characteristics of the respondents and the context as these should be representative of the population and the actual conditions that the findings will be generalised and applied to.

3.1.9. Social desirability bias

Studies have also advocated for the inclusion of a social desirability scale to assess for response bias and to ensure reliability of the tool. Evaluation of whether the tool assesses and mitigates social desirability bias or not should be done. This is because respondents tend to answer questions in a way they believe is acceptable, and which makes them to be viewed favourably by others [35].

3.1.10. Internal consistency

Internal consistency was also found to be an important measure for reliability, with the Cronbach Alpha coefficient described as the most appropriate measure when determining the internal consistency of assessment tools (Fig. 1). Findings also showed that the magnitude of the coefficient should be considered as it helps evaluate the degree to which different test items can produce similar

Table 1
An evaluation of commercial tools.

Component	Focus area	Criteria	Tool				
			CT1	CT2	CT3	CT4	CT5
1. Content Validity	Built on existing science	Tool built on existing science	✓	✓	✓	✓	×
		Defined approach used for content analysis	✓	✓	✓	✓	×
	Expert validation	Tool evaluated by experts	✓	✓	✓	✓	✓
		Checks done to assess whether the tool measures what it is intended to measure	✓	✓	×	✓	×
		Defined approaches used to quantify the judgement of content validity	×	✓	×	×	×
		The tool shows that it fairly and comprehensively covers the items or domain that it purports to cover	✓	✓	✓	×	×
		The questionnaire items were sufficient to measure the food safety culture of an organisation	✓	✓	×	✓	×
	Weighting	Weight factors were added to the dimensions	×	×	×	×	×
2. Face validation	Expert validation	The tool was assessed by food safety culture experts to determine the relevance of the tool to assess food safety culture	✓	✓	✓	✓	×
		Checks done to establish whether questions are relevant to assess food safety culture	✓	✓	×	✓	×
		Checks where done to evaluate whether the tool is easily understandable	✓	✓	×	✓	×
	Preliminary pilot testing	Preliminary pilot testing to check for effectiveness of the tool	✓	✓	×	✓	×
3. Construct validity	Acceptability	The construct is acceptable to measure food safety culture	✓	✓	✓	✓	×
	Correlation	The scale of the tool correlates with existing tools that measure the same construct	✓	✓	✓	✓	✓
		The correlation co-efficient is mentioned	×	×	×	×	×
4. Cultural Validity	Ethics	Research tools ethical and appropriate according to the standards of the target culture	✓	✓	×	✓	×
	Appropriateness	Tools and other information translated in a culturally appropriate way	✓	✓	✓	✓	×
5. Criterion validity	Predictive validity	The collected data highly correlates with data acquired in the first round of the research	–	–	–	–	–
		The tool predicts what is expected theoretically	✓	✓	×	✓	×
	Concurrent validity	The data correlates with data gathered from another tool i.e., agreement with a second measure	×	×	×	×	×
	Postdictive validity	Scores relate to the scores from an established tool or criterion administered at a previous point in time	×	×	×	×	×
6. Pilot testing	Relevance of tool	Pilot testing done to confirm the relevance of the tool to measure food safety culture	✓	✓	×	✓	×
7. Factor Analysis	Principal components	Principal components measured	×	✓	×	×	×
	Exploratory analysis	Exploratory factor analysis done	×	×	×	×	×
	Confirmatory analysis	Confirmatory factor analysis done	×	×	×	×	×
	KMO	The Kaiser-Meyer-Olkin value is stated	×	✓	×	×	×
8. Ecological validity	Generalisability	The tool can be generalised	✓	✓	✓	✓	×
9. Reliability	Social desirability	Social desirability bias	×	✓	×	×	×
	10.	Tests done to measure internal consistency	×	✓	×	×	×
	Internal consistency	Correlation of the questionnaire items	×	✓	×	×	×
	Other reliability checks	Other tests done to ensure reliability of the tool	✓	✓	✓	✓	×
		Statistical analysis done stated for each of the mentioned tests	×	✓	×	✓	×
		Statistical findings given for other reliability tests mentioned	×	✓	×	✓	×

✓means that the validity or reliability checks were done.
×means that there was no available information showing that the reliability or validity check had been done.

results when testing the same construct. Checking for both the type of internal consistency test done and the correlation co-efficient could enable authentication of the tool validation process.

3.1.11. Repeatability

Repeatability is also considered important as measurement errors can be found in differences across raters, different respondents and in content sampling [22]. The method used to check for repeatability should therefore be checked when evaluating tools for repeatability. Several methods have been identified including inter-rater reliability and test-retest reliability.

3.2. Evaluation of assessment tools

Table 1 shows findings from the evaluation of commercial food safety culture assessment tools against the assessment methodology. A total of five tools were identified and evaluated. CT2 was the most valid tool with CT5 being the least valid. Content (step 1.1) and face validity (step 1.2), and pilot testing (step 1.6) were the most evaluated with at least three out of the five tools (CT1, CT2 and CT4), assessing most criteria of these factors. Factor analysis (step 1.7) and reliability (steps 2.1–2.3) were the least evaluated. Ecological validity (step 1.8) was also checked for at least four out of the five commercial tools. CT2 was the only tool were factor analysis and social desirability bias checks were done. Interestingly none of the tools assigned weights to the food safety culture dimensions/ factors/elements as revealed by the content validation. Additionally, none of the tools were assessed for concurrent and postdictive validity, and the correlation coefficient relating to construct validity, with the predictive validation only measuring only one criterion i.e., whether the tool predicts what is expected theoretically and not demonstrating how the collected data highly correlates with data acquired in the first round of the research.

Table 2 shows the analysis of scientific tools. ST2 was the most validated tool with ST1 being the least validated. Like the

Table 2
An evaluation of the scientific tools.

Component	Focus area	Question (s)	Tool		
			ST1	ST2	ST3
1. Content Validity	Built on existing science	Tool built on existing science	✓	✓	✓
		Defined approach used for content analysis	✓	✓	✓
	Expert validation	Tool evaluated by experts	×	✓	×
		Checks done to assess whether the tool measures what it is intended to measure	×	✓	×
		Defined approaches used to quantify the judgement of content validity	×	✓	×
		Tool shows that it fairly and comprehensively covers the items or domain that it purports to cover	×	✓	×
		The questionnaire items were sufficient to measure the food safety culture of an organisation	×	✓	×
	Weighting	Weight factors were added to the dimensions	×	×	×
2. Face validation	Expert validation	Tool assessed by food safety culture experts to determine the relevance of the tool to assess food safety culture	✓	✓	✓
		Checks done to establish whether questions are relevant to assess food safety culture	×	✓	✓
		Checks where done to evaluate whether the tool is easily understandable	✓	✓	✓
	Preliminary pilot testing	Preliminary pilot testing to check for effectiveness of tool based on the respondent's understanding of the questions	✓	×	✓
3. Construct validity	Acceptability	The construct is acceptable to measure food safety culture	✓	✓	✓
	Correlation	The scale of the tool correlates with existing tools that measure the same construct	✓	✓	✓
		The correlation co-efficient is mentioned	×	×	×
4. Cultural Validity	Ethics	Research tools ethical and appropriate according to the standards of the target culture	×	×	✓
	Appropriateness	Tools and other information translated in a culturally appropriate way	×	✓	✓
5. Criterion validity	Predictive validity	The collected data highly correlates with data acquired in the first round of the research	–	–	–
		The tool predicts what is expected theoretically	✓	✓	✓
	Concurrent validity	The data correlates with data gathered from another tool i.e., agreement with a second measure	×	×	×
	Postdictive validity	Scores relate to the scores from an established tool or criterion administered at a previous point in time	×	×	×
6. Pilot testing	Relevance of tool	Pilot testing done to confirm the relevance of the tool to measure food safety culture	✓	✓	✓
7. Factor Analysis	Principal components	Principal components measured	×	×	✓
	KMO	The Kaiser-Meyer-Olkin value obtained is stated	✓	✓	✓
	Exploratory analysis	Exploratory factor analysis done	✓	✓	×
	Confirmatory analysis	Confirmatory factor analysis done	×	×	×
Ecological validity	Generalisability	The tool can be generalised	×	×	✓
8. Reliability	Social desirability	Check for social desirability bias	×	×	×
	9.	Tests done to measure internal consistency	✓	✓	×
	Internal consistency	Correlation of the questionnaire items	✓	✓	×
	Other reliability checks	Other tests done to ensure reliability of the tool	✓	×	✓
		Statistical analysis stated for each of the mentioned tests	×	×	×
		Statistical findings given for other reliability tests mentioned	×	×	×

commercial tools, food safety culture elements/factors/dimensions of the scientific tools were equally weighted as the findings show that weights were not assigned. Similar findings were also seen for predictive, concurrent, postdictive and construct validity. None of the scientific tools were evaluated for social desirability bias and no statistical analysis were done for the reliability tests. Face and criterion validation were the most validated in addition to the pilot testing and the factor analysis. Content, ecological, and cultural validity were the least evaluated.

Further analysis (Table 3) showed that CT2 scored highest when compared to the other tools i.e., both commercial and scientific tools as published peer reviewed scientific papers, demonstrated the tool's validity and reliability. CT5 was scored lowest as there was limited information on the tool. A comparison of both commercial and scientific tools shows that for at least three commercial tools 50 % of the validation criteria were covered by the tools i.e., CT1 (54.5 %), CT2 (78.8 %), and CT4 (51.5 %), when compared to the scientific tools were only ST2 covered 57.6 % of the validation criteria. However, the lowest scoring tools were also commercial tools with CT5 only covering 9.4 %.

4. Discussion

This research sought to evaluate the validity and reliability of food safety culture assessment tools. To achieve this a science-based method was developed to evaluate the assessment tools followed by the actual evaluation of both commercial and scientific food safety culture assessment tools. Jespersen, Griffiths, and Wallace (2017) suggested that having an established science-based approach provides a way to determine the trustworthiness of results against accepted methods. Eleven factors were therefore identified through a review of existing scientific principles and were deemed essential to evaluate the validity and reliability of food safety culture assessment tools.

4.1. Validity

4.1.1. Content validity

To demonstrate content validation, tool evaluations should consider whether the tool is based on existing theories. At least six out of eight tools were based on existing theories, whilst one (CT3) was based on industry guidance documents and tools (Table 1, Table 2). Almanasreh, Moles, and Chen [36] suggested that tools should be theory based. Existing food safety culture studies have also explored literature from other cultural domains [37,38], human dimensions [6,39–41], education [42], management science [42], and social science and food science [43] in their tool development process. The approach used for content analysis should also be specified. Examples include content validity ratios, content validation forms, deductive content analysis and iterative approaches [15]. The evaluated tools used deductive analysis (CT2, ST1), inductive analysis (ST1), and iterative (CT1, ST3) approaches. Additionally, food safety culture experts such as technical experts, academics, and industry professionals should validate the content. Experts from other relevant domains such as social scientists can also be part of the expert panel. Kirezieva et al. [44] suggested that experts in the field of relevance need to evaluate the reliability, relevance, and validity of the tools. All commercial tool developers ensured that experts (Table 1) had evaluated their tools, and only ST2 for the scientific tools. Content validation also depends on the effectiveness of the expert panel [15]. None of the tool descriptions discussed criteria used to evaluate effectiveness of the panel. Quatrini Carvalho Passos Guimarães et al. (2016) suggested that the adoption of inappropriate criteria for selection of the experts required to validate tools could affect the tool validity and the subsequent findings Studies often suggest experts within their academic institutions ([6];Spagnoli, Jacxsens, & Vlerick, 2023) and selection criteria of these experts is not published. Utilizing selection criteria where expert criteria is defined and an expert scoring provided [45], and different experts are considered, including users of the tool could be considered in the tool development process.

To ensure valid assessments, how the different constructs differently impact an organisation's prevailing food safety culture and maturity should be evaluated (Zanin, Stedefeldt, & Luning, 2021). For both the commercial and scientific tools, the elements/dimensions/factors were equally weighted, and it was not clear how these differently contribute to an organisation's food safety culture and culture maturity, and ultimately food safety performance. Adding weight factors could help understand the different contributions

Table 3
Overall score for validity and reliability of the tools.

Organisation	Commercial tools					Scientific Tools		
Tool	CT1	CT2	CT3	CT4	CT5	ST1	ST2	ST3
Component								
1. Content Validity	6	7	4	6	1	2	7	2
2. Face validation	4	4	2	3	1	3	3	4
3. Construct validity	2	2	2	2	1	2	2	2
4. Cultural Validity	2	2	1	2	–	–	1	2
5. Criterion validity	1	1	–	1	–	1	1	1
6. Pilot testing	1	1	–	1	–	1	1	1
7. Factor Analysis	–	2	–	–	–	2	2	2
8. Ecological validity	1	1	1	1	–	–	–	1
9. Reliability	1	6	1	1	–	3	2	1
Total Score (33)	**18**	**26**	**11**	**17**	**3**	**14**	**19**	**16**

of the constructs [46]. Most studies recommend adding weight factors ([47]; Zanin, Stedefeldt, & Luning, 2021) with limited studies assigning these weights [48].

4.1.2. Face validity

Tool evaluation should consider whether experts in the field of study have been used to achieve face validity [49]. Our findings showed that a majority (3/5) of the commercial tools and all the scientific tools had undergone face validation (Table 1, Table 2). Technical and subject matter experts evaluated the tools. Peer review also contributed to face validation. Relevance, context, confirmability, terminology, adequacy, dependability, consistency, intelligibility, completeness, ordering of questions, answer options, facilitation of honest responses, applicability, practicality, clarity, meaningfulness, formatting, and grammar were mostly checked. Findings are consistent with Oluwatayo [49] who suggested that face validity is done to check whether the tool is understandable, including readability, feasibility, clarity, and to confirm whether the format and presentation are relevant to be considered as a measuring tool, i.e., whether the tool is reasonable, clear, and unambiguous to understand and respond. Cotter, Yamamoto, and Stevenson [50] indicated that although face validation should include experts and the target audience, time and money constraints, and lack of available experts often results in lack of validation.

Preliminary pilot testing should also be evaluated to judge the effectiveness of the questionnaire before the fully fledged pilot test [15]. This provides an opportunity for the tool developers to have insights into items where the respondents have confusion and where there are suggestions for improvements [22]. Findings show that three commercial tools and two scientific tools (Table 1, Table 2) were preliminarily pilot tested. Pretesting is crucial as it detects problems with the tool, reducing respondent burden and ensuring that the questions do not influence respondents reactions [51].

4.1.3. Construct validity

To demonstrate construct validity, the constructs should be acceptable to measure food safety culture. Both commercial and scientific tools, except for CT5 used accepted dimensions/factors/elements that permeate food safety culture literature [52–54]. Mechanisms should also be in place to confirm the tool measures what it is intended to measure. Evaluation of the tools showed that peer review, management in assessed companies, respondents, and subject matter experts (Table 2) agreed that the tools were an acceptable measure for food safety culture. Additionally, whether the scale of the tool correlates with existing tools that measure the same construct should be evaluated. All evaluated tools used common assessment methods and scales, and these resonated with existing studies. For example, most tools used a five-point Likert scale and stages of maturity. Several tools used in researching food safety culture also use a Likert scale [8,39,42,55–57] and stages of maturity [9,43,50,58] to gain insight into an organisation's food safety culture. However, the correlation coefficient was not considered, and this could be achieved by administering the tool with an existing tool measuring similar constructs to a similar group of respondents and estimating the associations by developing correlation matrices [15]. The analysis could check for either convergent or discriminant validity. Jespersen, MacLaurin, and Vlerick [59] suggested that variations caused by the data collection method could be a threat to construct validity and suggested the use of Pearson product-moment correlations and ANOVA to measure construct and discriminant validity in food safety culture assessments.

Furthermore the tool should measure enduring characteristics of the organisation if it purports to assess the culture of the organisation and this can be done through repeated assessments over time and phrasing of questions or surveys, which assess the enduring characteristics rather than the climate [31,54].

4.1.4. Cultural validity

To ensure cultural validity, the tool needs to be "grounded" in that culture if it is to be considered valid [32]. It is postulated that culture and society shape an individual's values, beliefs, experiences, mind and thinking and to attain cultural validity, the tool needs to be adapted to the context of the respondents [60]. The tool should be tailored to the context. Three of the five commercial tools were customised, which made the tools appropriate for the targeted respondents. However, details on any ethical considerations were not elaborated as only one of the scientific tools discussed the ethical approvals done. To ensure the tool is appropriate for the context it is important to translate the tools in a culturally appropriate way. Four commercial and two scientific tools were translated into the language of respondents and one tool considered the culture of the respondents. Findings are consistent with Joomun et al. [54], who also translated the food safety culture questionnaire using food safety experts to ensure correct translation and cultural equivalence. However, translation only contributes to the validity when appropriately done. Shi, Pinto, and Guanais [61]suggested that translation goes beyond language translation but should ensure a holistic, rigorous transcultural process which considers the feasibility of the tool in different cultural contexts. To ensure cross cultural adaptation, a six-steps process has been recommended to ensure valid cultural and linguistic translations i.e., forward translation, qualitative review, backward translation, joint review by experts and translators, lay panel review and pilot testing [62].

Additionally, whilst food safety culture research acknowledges the role of national culture in shaping the food safety culture of food businesses [63–65], there is limited research on this association. Understanding the national values could be useful in ensuring that the tools are designed and translated in a culturally appropriate way and ensure they are appropriate to the target culture. Nyarugwe et al. [37] suggested that incorporating and understanding these national values could enable the appropriate research tools to be used as an approach considered effective in one culture could be perceived differently in another due to differences in predispositions, and food safety risk perceptions. Experts in the national culture domain suggest that adapting to the national culture is key to ensuring cultural validity [66,67].

4.1.5. Criterion validity

Assessment tools should demonstrate the extent to which results from one measure can be associated with results from another measure [68,69]. One way is to ensure predictive validity, where the assessment tool should predict future outcomes or behaviour and in the context of food safety, assess the ability of a tool to predict future food safety performance or incidences. Food safety performance should be assessed over time by conducting longitudinal studies. By comparing the food safety outcomes with the food safety culture assessment scores, the extent to which food safety culture assessments predict food safety performance would give an indication of the predictive validity of the tool. Evaluations of predictive validity should also check whether the data collected highly correlates with data acquired in the first round of the research and whether the tool should predict what is expected theoretically. All scientific tools and only three commercial tools (CT1, CT2, CT4) predicted what is expected theoretically. This is because findings were corroborated by other studies, literature, and existing theories. Jespersen, Griffiths, and Wallace (2017) also found out that a few tools showed predictive validation as studies focused on face and construct validity. However, correlations with previous data were not checked. A lack of predictive validity has been seen as a major hurdle in assessment tools [70].

Another way to establish criterion validity is by concurrent validation, which evaluates the operationalizability of the tool and the level of agreement between two different assessments (Granpeesheh et al., 2014 [21]). As such, how well the food safety culture tool correlates with an established measure of food safety culture should be evaluated as this demonstrates the tool's ability to produce results that are comparable to an established measure. However, Postdictive validity can also be useful to demonstrate the ability of a tool to accurately predict past events (retrospective validity) and evaluations should check whether the tool can explain historical food safety outcomes/performance. None of the tools demonstrated concurrent and postdictive validity, based on the available data.

4.1.6. Pilot testing

The final version of the questionnaire should mirror the full study and be tested with a considerable number of the intended respondents to confirm the relevance of the tool [15]. Findings show that only three commercial tools (CT1, CT2 and CT4) were pilot tested and interestingly, these had also been preliminary pilot tested. Cotter, Yamamoto, and Stevenson [50] found only 22 % studies included pilot testing. In (2017) [71] indicated that despite the importance and benefits of conducting a pilot study, researchers are often not interested and that many are not reported as most focus on statistical outcomes rather than feasibility of the study. All scientific tools were pilot tested supporting the evidence that this step is crucial in ensuring the feasibility of the study.

4.1.7. Factor analysis

Factor analysis is useful to explore the underlying dimensions/factors/elements valuable in understanding the food safety culture concept. It is therefore a key technique to consider when evaluating the validity of an assessment tool. The technique is useful in examining the underlying structure and can be utilised to demonstrate construct validity, to evaluate the number of dimensions sufficient in explaining the variance in observed variables, to remove redundant factors, and to demonstrate internal consistency and the stability and generalisability to ensure the tool measures consistently across different groups [72,73]. Only one commercial tool (CT2) was subjected to factor analysis and all three scientific tools had the analysis done. However, the type of analysis differed with two scientific tools using EFA, one scientific and one commercial tool using PCA and ST1 using principal axis analysis. EFA is one of the main types of factor analysis, which is often done in the early phases of research to obtain information on the interrelationships between a set of variables to check for dimensionality [74]. It also helps in exploring the underlying theoretical structure and underlying factors that should be retained i.e., the minimum number of common factors required to adequately reproduce the item correlation matrix [75].

CFA is a more complex technique that tests for specific hypotheses or theories concerning the structure underlying a set of variables [76]. It evaluates how well the measured variables represent the number of constructs, and is useful in determining the "relationships between observed measures or indicators (e.g., test items, test scores, behavioural observation ratings) and latent variables to establish the number and nature of factors that account for the variation and covariation among a set of indicators [77]. Izquierdo Alfaro, Olea Díaz, and Abad García (2014) suggested that it is good practice to conduct a cross-validation applying both EFA and CFA. PCA is used as a dimensionality reduction technique. Several studies including Gruijters [78] discuss the inadequacies of PCA as it mainly achieves data reduction and indicate that factor analysis has a broader context which also encompasses latent variable analysis and analysis of interrelationships. However, PCA with the varimax rotation method has been found useful in validating both convergent and divergent construct validity [15,21]. As these are statistical techniques, tool evaluations should also consider the statistical analysis and findings (e.g., factor loadings, eigenvalues, factor correlation coefficients) that help in determining the validity of the tools. The Kaiser-Meyer-Olkin (KMO) is a common measure, which demonstrates how suited the data is for factor analysis [76] and appeared to be commonly assessed by the tools evaluated. The KMO value obtained for the scientific tools ranged from 0.67 to 0.93. For CT2, the KMO was greater than 0.7. Values greater than 0.6 are often acceptable although values between 0.7 and 0.79 are often described as middling and 0.6–0.69 as mediocre [76,79].

4.1.8. Ecological validity

It is important that the tool has real-world application and reflect real world contextual factors, dynamics, and conditions. Most commercial tools (Table 1) had real world application in multiple countries, and companies, differing in operational characteristics such as size, sector, and riskiness, and in participants from distinct roles and positions within the organisation. A study by Spagnoli, Jacxsens, and Vlerick [40] suggested that an approach that utilises on-site evidence collection, and considers repeated food safety culture assessments in different locations within the factory, and different shifts demonstrates ecological validity. Although this is important in obtaining naturalistic data, the context should be considered as it affects the generalisability of a tool [80]. The context encompasses characteristics of a company's environment e.g., food safety governance, national values, internal company characteristics such as vulnerability and characteristics of the products, process, chain environment as described by Kirezieva et al. [33].

4.2. Reliability

Reliability refers to the accuracy and consistency of test scores [81,82]. Evaluating the reliability of an assessment tool helps to check for errors present in content sampling and measurement scales. Commonly used criteria include checks for internal consistency, social desirability bias, and other reliability tests such as inter-rater reliability.

4.2.1. Internal consistency

When evaluating the reliability or consistency of questions within an assessment tool, the Cronbach's alpha provides a measure of whether the questionnaire items consistently assess the same underlying construct [21,76]. Only one commercial tool (CT2) and two scientific tools (ST1 and ST2) assessed internal consistency using Cronbach's alpha. The Cronbach's alpha was found to be between 0.5 and 0.83, 0.668–0.922, and 0.97 for CT2, ST1 and ST2, respectively. A high level of the Cronbach's alpha value shows a high internal consistency, which demonstrates the questionnaire items are highly correlated and measure the same construct consistently. A value of more than 0.7 is acceptable [76]. Jenkinson, Wright, and Coulter [68] suggested that values > 0.5 are acceptable. However, low scores suggest that the items may not be measuring the construct reliably.

4.2.2. Social desirability bias

To ensure reliability, social desirability should be assessed as food safety culture tools often involve self-reported measures. Due to various issues e.g., collecting personal data, data on questions relating to widely accepted attitudes or behavioural/social norms, recall of incidents, a social desirability scale can be added to the questionnaire allowing the researcher to assess the extent to which social desirability bias influences the survey responses [35]. Hinsz, Nickell, and Park [83] highlighted the issues with self-reporting as employees are likely to reflect their behaviours positively i.e., in a socially desirable manner and underreport undesirable behaviour and supported the idea of embedding a socially desirable measure in the assessment tools. In our study only CT2, included a social desirability scale. Other studies [54,63] have also considered the importance of including a social desirability scale and approaches to mitigate social desirability bias including confidential surveys, participant observations, balancing positive and negative statements and validating self-reported data using other qualitative methods (method-triangulation) have been proposed.

4.2.3. Other reliability assessments

In addition to internal consistency and social desirability bias, other reliability assessments can be used to demonstrate the extent to which the tool produces consistent and dependable results when administered under different conditions. Examples include test-retest, inter-rater and split half reliability which measure the stability of the tool over time, consistency of ratings between raters and the extent to which two halves of the tool yield consistent results, respectively. For CT2, inter-rater reliability was assessed through pairwise agreement, and checks were done for discrimination between food safety culture dimensions through cluster analysis. A 90.4 % pairwise agreement and Pearson correlation values at or equal to 0.5, or less than 0.5 for similar and different items, were obtained for the pairwise agreement and cluster analysis, respectively. Jespersen and Wallace [84] discussed the importance of pairwise agreement and suggested a level of 90 % to be appropriate in food safety culture assessments. For CT3 an assessment team with complimentary skills and expertise assessed the tool. This is consistent across food safety culture tools development approaches where the process is often iterative [6,38,40]. However, no statistical analysis was done to assess consistency between the raters. Similarly, two professionals reviewed the findings in ST1, but the reliability score was not given. A within-methods triangulation was done for ST3. Cooper [85] indicated that a within-method validation can be used in checking for internal consistency and reliability.

5. Conclusion

The study confirmed our hypotheses and established that having an established science-based approach provides a way to determine the trustworthiness of established assessment tools against science-based and accepted methods. Moreover, demonstrating rigor of the tool development process is important as this ensures transparency, credibility, and usefulness of research and practical findings to avoid misleading those who use the outcomes. This is because results and inferences made from assessments based on these tools are impacted by the validity and reliability of the tool. Our findings show that validation of each of the eight assessment tools focuses on specific aspects as some areas were not validated. Moreover, there is a misconception that if a tool is applied in actual practice, and validated by experts, then it has been validated. However, for a tool to be considered validated, it should undergo a rigorous process, which considers the criteria that have been proposed. An established science-based approach therefore provides a useful analysis or check for assessment tools thus ably guiding regulators, the food industry and academia in their approach to evaluating food safety culture assessment tools and could provide a starting point to which a reflection of tools can be done. Strengths and weaknesses can therefore be identified, which can enable improvements of the tools to enhance the accuracy of the assessments or for an organisation to determine if the tool is the right fit. To enable an organisation to determine if the tool is the right fit, it is important that information pertaining to these validation checks is made available. Moreover, credible data can be valuable in identifying improvement areas and implementing targeted interventions.

It is also important to understand that validity of a tool can be influenced by other factors including context, the characteristics of the organisation where the tool is applied, the tool used, and the respondents. The constructs, validation checks, statistical analysis and interpretation thereof require careful consideration and expertise. Having these insights is crucial when developing the tool and in interpreting findings.

A limitation of this study was the restricted information on commercial tools as this was not publicly available and attempts to obtain the information from the tool developers were futile. Transparency and information sharing could allow more tools to be evaluated enabling clear recommendations and better use of assessment results. It is also important to acknowledge that lack of detailed descriptions on how assessment tools were validated may result in methodological inaccuracies. Moreover, flaws in the assessment tool design, food safety culture evaluation and analysis of findings could impact not only the validity and reliability of the tool but also mislead those who are investing in improving their culture of food safety. Particularly, acknowledging the cultural biases and challenges posed by translation of the tools and how this affects the applicability of the tool across different contexts is important. Tool developers should therefore avoid direct translation of the tools, but apply standardized guidelines for translation and adaptation e.g., Johns Hopkins recommendations [61], followed by a psychometric validation method to ensure a valid and reliable assessment tool for the specific country's context.

The importance of reliable and valid food safety culture assessment tools is increasingly being recognised as food businesses acknowledge the strategic importance of food safety culture to their businesses and the cost of poor quality and inadequate food safety [41,86]. The need for validated assessment tools has never been more important as data from these assessment tools guides judgements, decision-making and interventions. Further research should evaluate the role of data in maturing culture and further explore the economic implications.

Further research will also explore ecological validity to gain insights from practitioners' perspectives to enhance practical relevance. Additionally, future research should explore active approaches and digitally enabled teaching and learning approaches to educate future food safety professionals [87,88]. This is important as such strategies could be beneficial in producing well-rounded practitioners who would not only be proactive, but also be able to address challenges with food safety culture assessments, challenge assumptions, and establish the field as has been done in other culture domains.

CRediT authorship contribution statement

Shingai P. Nyarugwe: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Lone Jespersen:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:The authors work with one of the commercial research tools, which is part of the analysis. However, this did not influence the work reported in this paper and the information on this tool is available in the public domain.

Acknowledgements

We are grateful to the industry practitioners and the academicians who assisted with the identification of the commercial products, and the commercial tool developers who assisted with information on their tools.

References

[1] European Commission, COMMISSION REGULATION (EU) 2021/382 of 3 March 2021 Amending the Annexes to Regulation (EC) No 852/2004 of the European Parliament and of the Council on the Hygiene of Foodstuffs as Regards Food Allergen Management, Redistribution of Food and Food Safety Culture, Official Journal of the European Union, 2021.
[2] FSANZ, Food safety culture. https://www.foodstandards.gov.au/foodsafety/culture/Pages/default.aspx, 2023.
[3] FDA, NEW ERA of SMARTER FOOD SAFETY FDA's Blueprint for the Future, U.S. Food and Drug Administration, Maryland, USA, 2020.
[4] GFSI, GFSI Benchmarking Requirements, 2020. Retrieved from, https://mygfsi.com/wp-content/uploads/2020/02/GFSI-One-pager-Benchmarking-Requirements-v2020-vWeb-1.pdf.
[5] B. Ball, A. Wilcock, S. Colwell, Tool for measuring food safety climate, J. Food Protect. 73 (2010) 84.
[6] E. De Boeck, L. Jacxsens, M. Bollaerts, P. Vlerick, Food safety climate in food processing organizations: development and validation of a self-assessment tool, Trends Food Sci. Technol. 46 (2) (2015) 242–251.
[7] U.Z.A.U. Fatimah, S.W. Arendt, C.H. Strohbehn, Food safety culture in onsite foodservices: development and validation of a measurement scale, Journal of Foodservice Management and Education 8 (1) (2014) 1.
[8] S.T. Wu, S.R. Hammons, R. Silver, J.A. Neal, H.F. Oliver, Retail deli managers and associates have better food safety culture in stores with lower Listeria monocytogenes contamination, Food Control 110 (2020) 106983.
[9] L.M. Zanin, P.A. Luning, D.T. da Cunha, E. Stedefeldt, Influence of educational actions on transitioning of food safety culture in a food service context: Part 1 – triangulation and data interpretation of food safety culture elements, Food Control 119 (2021) 107447, https://doi.org/10.1016/j.foodcont.2020.107447.
[10] L.M. Zanin, E. Stedefeldt, P.A. Luning, The evolvement of food safety culture assessment: a mixed-methods systematic review, Trends Food Sci. Technol. 118 (2021) 125–142.
[11] Food and Agricultural Organization, & World Health Organization, in: Guidelines for the Validation of Food Safety Control Measures. CAC/GL 69 - 2008, FAO/WHO joint commission, Rome, 2013 (Vol. CAC/GL 69 - 2008).
[12] Food and Agricultural Organization, & World Health Organization, Guidelines on performance criteria and validation of methods for detection, identification and quantification of specific DNA sequences and specific proteins in foods, in: CXG 74-2010 in (Vol. CXG 74-2010), FAO/WHO joint commission, Rome, 2010.
[13] Food and Agricultural Organization, & World Health Organization, in: Harmonized IUPAC Guidelines for Single-Laboratory Validation of Methods of Analysis. CAC/GL 49-2003, FAO/WHO joint commission, Rome, 2003 (Vol. CAC/GL 49-2003).
[14] Food and Agricultural Organization, & World Health Organization, General principles of food hygiene CXC 1-1969, in: Codex Alimentarius International Food Safety Standards, CXC vols. 1–1969, FAO/WHO joint commission, Rome, 2022.

[15] A. Aithal, P. Aithal, Development and validation of survey questionnaire & experimental data–A systematical review-based statistical approach, International Journal of Management, Technology, and Social Sciences (IJMTS) 5 (2) (2020) 233–251.
[16] W.M.K. Trochim, Reliability and validity, Research Methods Knowledge Base (2006), 2006. https://conjointly.com/kb/reliability-and-validity/. (Accessed 22 May 2023). Retrieved from.
[17] P. Roberts, H. Priest, Reliability and validity in research, Nurs. Stand. 20 (44) (2006) 41–46.
[18] W.M. Trochim, J.P. Donnelly, Research methods knowledge base, in: Atomic Dog Pub, vol. 2, Macmillan Publishing Company, New York, 2001.
[19] L. Cohen, L. Manion, K. Morrison, Research Methods in Education, sixth ed., Routledge, United Kingdom, 2007.
[20] K. Fitzner, Reliability and validity A quick review, Diabetes Educat. 33 (5) (2007) 775–780, https://doi.org/10.1177/0145721707308172.
[21] H. Taherdoost, Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research, in: How to Test the Validation of a Questionnaire/survey in a Research, 2016 *(August 10, 2016)*.
[22] S. Tsang, C.F. Royse, A.S. Terkawi, Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine, Saudi J. Anaesth. 11 (5) (2017) 80.
[23] L. Jespersen, M. Griffiths, C.A. Wallace, Comparative analysis of existing food safety culture evaluation systems, Food Control 79 (2017) 371–379.
[24] G.M. Tawfik, K.A.S. Dila, M.Y.F. Mohamed, D.N.H. Tam, N.D. Kien, A.M. Ahmed, N.T. Huy, A step by step guide for conducting a systematic review and meta-analysis with simulation data, Trop. Med. Health 47 (1) (2019) 1–9.
[25] A. Grewal, H. Kataria, I. Dhawan, Literature search for research planning and identification of research problem, Indian J. Anaesth. 60 (9) (2016) 635–639.
[26] L.A. Palinkas, S.M. Horwitz, C.A. Green, J.P. Wisdom, N. Duan, K. Hoagwood, Purposeful sampling for qualitative data collection and analysis in mixed method implementation research, Administration and policy in mental health 42 (5) (2015) 533–544, https://doi.org/10.1007/s10488-013-0528-y.
[27] S. Tiku, M. Pecht, Validation of reliability capability evaluation model using a quantitative assessment process, Int. J. Qual. Reliab. Manag. 27 (8) (2010) 938–952.
[28] O. Isaksson, C. Eckert, M. Panarotto, J. Malmqvist, You need to focus to validate. Proceedings of the Design Society: DESIGN Conference, 2020. Paper presented at the.
[29] K. Kirezieva, L. Jacxsens, M. Uyttendaele, M.A. Van Boekel, P.A. Luning, Assessment of food safety management systems in the global fresh produce chain, Food Res. Int. 52 (1) (2013) 230–242.
[30] R. Botreau, I. Veissier, A. Butterworth, M.B. Bracke, L.J. Keeling, Definition of criteria for overall assessment of animal welfare, Anim. Welf. 16 (2) (2007) 225–228.
[31] D.A. Wiegmann, H. Zhang, T.L. Von Thaden, G. Sharma, A.M. Gibbons, Safety culture: an integrative review, Int. J. Aviat. Psychol. 14 (2) (2004) 117–134.
[32] D. Jadhav, What is cultural validity and why is it ignored?, in: AMB Diemen 2009, 2009.
[33] K. Kirezieva, L. Jacxsens, G.J. Hagelaar, M.A. van Boekel, M. Uyttendaele, P.A. Luning, Exploring the influence of context on food safety management: case studies of leafy greens production in Europe, Food Pol. 51 (2015) 158–170.
[34] J. DeCoster, Overview of Factor Analysis, 1998.
[35] P. Grimm, Social desirability bias, 2, Wiley international encyclopedia of marketing, New Jersey, 2010, p. 258.
[36] E. Almanasreh, R. Moles, T.F. Chen, Evaluation of methods used for estimating content validity, Res. Soc. Adm. Pharm. 15 (2) (2019) 214–221.
[37] S.P. Nyarugwe, A. Linnemann, G.J. Hofstede, V. Fogliano, P.A. Luning, Determinants for conducting food safety culture research, Trends Food Sci. Technol. 56 (2016) 77–87.
[38] W.S. Alrobaish, L. Jacxsens, P. Spagnoli, P. Vlerick, Assessment of food integrity culture in food businesses through method triangulation, Food Control 141 (2022) 109168.
[39] E. De Boeck, A.V. Mortier, L. Jacxsens, L. Dequidt, P. Vlerick, Towards an extended food safety culture model: studying the moderating role of burnout and jobstress, the mediating role of food safety knowledge and motivation in the relation between food safety climate and food safety behavior, Trends Food Sci. Technol. 62 (2017) 202–214, https://doi.org/10.1016/j.tifs.2017.01.004.
[40] P. Spagnoli, L. Jacxsens, P. Vlerick, Towards a food safety culture improvement roadmap: diagnosis and gap analysis through a conceptual framework as the first steps, Food Control 145 (2023) 109398.
[41] P. Spagnoli, L. Defalchidu, P. Vlerick, L. Jacxsens, The relationship between food safety culture maturity and cost of quality: an empirical pilot study in the food industry, Foods 13 (4) (2024) 571.
[42] J. Taylor, J.P. Garat, S. Simreen, G. Sarieddine, An industry perspective: a new model of food safety culture excellence and the impact of audit on food safety standards, Worldwide Hospitality and Tourism Themes 7 (1) (2015) 78–89.
[43] L. Jespersen, M. Griffiths, T. Maclaurin, B. Chapman, C.A. Wallace, Measurement of food safety culture using survey and maturity profiling tools, Food Control 66 (2016) 174–182.
[44] K. Kirezieva, J. Nanyunja, L. Jacxsens, J.G. van der Vorst, M. Uyttendaele, P.A. Luning, Context factors affecting design and operation of food safety management systems in the fresh produce chain, Trends Food Sci. Technol. 32 (2) (2013) 108–127.
[45] H.C. Quatrini Carvalho Passos Guimarães, S.B. Pena, J.d.L. Lopes, C.T. Lopes, A.L. Bottura Leite de Barros, Experts for validation studies in nursing: new proposal and selection criteria, International Journal of Nursing Knowledge 27 (3) (2016) 130–135.
[46] S.P. Nyarugwe, A. Linnemann, L.K. Nyanga, V. Fogliano, P.A. Luning, Food safety culture assessment using a comprehensive mixed-methods approach: a comparative study in dairy processing organisations in an emerging economy, Food Control 84 (2018) 186–196.
[47] L. Manning, The value of food safety culture to the hospitality industry, Worldwide Hospitality and Tourism Themes 10 (3) (2018) 284–296.
[48] J. Mu, Q. Li, Fuzzy comprehensive evaluation on food safety culture level in enterprise, in: Paper Presented at the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2015.
[49] J.A. Oluwatayo, Validity and reliability issues in educational research, Journal of educational and social research 2 (2) (2012), 391-391.
[50] S. Cotter, J. Yamamoto, C. Stevenson, A systematic characterization of food safety training interventions using the analyze, design, develop, implement, evaluate (ADDIE) instructional design framework, Food Control 145 (2023) 109415.
[51] B. Desye, A.H. Tesfaye, C. Daba, G. Berihun, Food safety knowledge, attitude, and practice of street food vendors and associated factors in low-and middle-income countries: a Systematic review and Meta-analysis, PLoS One 18 (7) (2023) e0287996.
[52] A. Kramer, E.R. Hoover, N. Hedeen, L. DiPrete, J. Tuttle, D. Irving, E. Moritz, Development of an empirically derived measure of food safety culture in restaurants, J. Food Protect. 86 (3) (2023) 100043.
[53] L.M. Zanin, P.A. Luning, E. Stedefeldt, A roadmap for developing educational actions using food safety culture assessment–A case of an institutional food service, Food Res. Int. 155 (2022) 111064.
[54] A.B. Joomun, C.A. Wallace, B. Aumjaud, D. Ramful-Baboolall, Food safety culture and climate prevailing in micro and small food manufacturing enterprises in Mauritius and Rodrigues, Int. J. Food Sci. Technol. 59 (2) (2024) 685–702.
[55] M.L. de Andrade, E. Stedefeldt, L.M. Zanin, D.T. da Cunha, Food safety culture in food services with different degrees of risk for foodborne diseases in Brazil, Food Control 112 (2020) 107152.
[56] U.Z.A.U. Fatimah, C.H. Strohbehn, S.W. Arendt, An empirical investigation of food safety culture in onsite foodservice operations, Food Control 46 (2014) 255–263.
[57] M. Wiśniewska, E. Czernyszewicz, A. Kałuża, The assessment of food safety culture in small franchise restaurant in Poland: the case study, British Food Journal 121 (10) (2019) 2365–2378.
[58] S.P. Nyarugwe, A.R. Linnemann, P.A. Luning, Prevailing food safety culture in companies operating in a transition economy-Does product riskiness matter? Food Control 107 (2020) 106803.
[59] L. Jespersen, T. MacLaurin, P. Vlerick, Development and validation of a scale to capture social desirability in food safety culture, Food Control 82 (2017) 42–47.
[60] G. Solano-Flores, S. Nelson-Barber, On the cultural validity of science assessments, J. Res. Sci. Teach.: The Official Journal of the National Association for Research in Science Teaching 38 (5) (2001) 553–573.

[61] L. Shi, D.M. Pinto, F.C. Guanais, Measurement of Primary Care: Report on the Johns Hopkins Primary Care Assessment Tool, 2013.
[62] N.T. Hoa, N.M. Tam, W. Peersman, A. Derese, J.F. Markuns, Development and validation of the Vietnamese primary care assessment tool, PLoS One 13 (1) (2018) e0191181.
[63] K. Fujisaki, M. Shimpo, R. Akamatsu, Factors related to food safety culture among school food handlers in Tokyo, Japan: a qualitative study, J. Foodserv. Bus. Res. 22 (1) (2019) 66–80.
[64] R. Nayak, P. Waterson, The Assessment of Food Safety Culture: an investigation of current challenges, barriers and future opportunities within the food industry, Food Control 73 (2017) 1114–1123.
[65] J. Taylor, An exploration of food safety culture in a multi-cultural environment: next steps? Worldwide Hospitality and Tourism Themes 3 (5) (2011) 455–466.
[66] G. Hofstede, G.J. Hofstede, M. Minkov, Cultures and Organizations: Software of the Mind: Intercultural Cooperation and its Importance for Survival, 2010.
[67] E. Meyer, Being the boss in Brussels, Boston, and beijing, Harv. Bus. Rev. 95 (4) (2017) 70–77.
[68] C. Jenkinson, L. Wright, A. Coulter, Criterion validity and reliability of the SF-36 in a population sample, Qual. Life Res. 3 (1) (1994) 7–12.
[69] J. Song, E. Howe, J.R. Oltmanns, A.J. Fisher, Examining the concurrent and predictive validity of single items in ecological momentary assessments, Assessment 30 (5) (2023) 1662–1671.
[70] K.N. Key, P.T. Hu, I. Choi, D.J. Schroeder, Safety culture assessment and continuous improvement in aviation: A literature review (DOT/FAA/AM-23/13), DC, Washington, 2023. https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/media/23-13-Safety%20Culture%20Assessment%20and%20Continuous%20Improvement%20in%20Aviation-%20A%20Literature%20Review_0.pdf. Retrieved from.
[71] J. In, Introduction of a pilot study, Korean journal of anesthesiology 70 (6) (2017) 601–605.
[72] R.F. DeVellis, C.T. Thorpe, Scale Development: Theory and Applications, Sage publications, 2021.
[73] R.L. Worthington, T.A. Whittaker, Scale development research: a content analysis and recommendations for best practices, Counsel. Psychol. 34 (6) (2006) 806–838.
[74] B. Williams, A. Onsman, T. Brown, Exploratory factor analysis: a five-step guide for novices, Australasian journal of paramedicine 8 (2010) 1–13.
[75] I. Izquierdo, J. Olea, F.J. Abad, Exploratory factor analysis in validation studies: uses and recommendations, Psychothema , 26 (3) (2014) 395–400.
[76] N. Shrestha, Factor analysis as a tool for survey analysis, Am. J. Appl. Math. Stat. 9 (1) (2021) 4–11.
[77] T.A. Brown, M.T. Moore, Confirmatory factor analysis, Handbook of structural equation modeling 361 (2012) 379.
[78] S. Gruijters, Using Principal Component Analysis to Validate Psychological Scales: Bad Statistical Habits We Should Have Broken Yesterday II, 2020.
[79] C.D. Dziuban, E.C. Shirkey, When is a correlation matrix appropriate for factor analysis? Some decision rules, Psychol. Bull. 81 (6) (1974) 358.
[80] M.A. Schmuckler, What is ecological validity? A dimensional analysis, Infancy 2 (4) (2001) 419–436.
[81] S.G. Sireci, The construct of content validity, Soc. Indicat. Res. 45 (1) (1998) 83–117.
[82] W.B. Schaufeli, S. Desart, H. De Witte, Burnout Assessment Tool (BAT)—development, validity, and reliability, Int. J. Environ. Res. Publ. Health 17 (24) (2020) 9495.
[83] V.B. Hinsz, G.S. Nickell, E.S. Park, The role of work habits in the motivation of food safety behaviors, J. Exp. Psychol. Appl. 13 (2) (2007) 105.
[84] L. Jespersen, C.A. Wallace, Triangulation and the importance of establishing valid methods for food safety culture evaluation, Food Res. Int. 100 (2017) 244–253.
[85] M.D. Cooper, Towards a model of safety culture, Saf. Sci. 36 (2) (2000) 111–136.
[86] L. Jespersen, J. Butts, G. Holler, J. Taylor, D. Harlan, M. Griffiths, C.A. Wallace, The impact of maturing food safety culture and a pathway to economic gain, Food Control 98 (2019) 367–379, https://doi.org/10.1016/j.foodcont.2018.11.041.
[87] L. Duedahl-Olesen, E. Holmfred, A.A. Niklas, I.K. Nielsen, J.J. Sloth, Case study teaching for active learning on analytical quality assurance concepts in relation to food safety exposure assessment, J. Chem. Educ. 98 (12) (2021) 3776–3783.
[88] E. Muca, D. Cavallini, F. Raspa, C. Bordin, D. Bergero, E. Valle, Integrating new learning methods into equine nutrition classrooms: the importance of students' perceptions, J. Equine Vet. Sci. 126 (2023) 104537.