School of Medicine, Nazarbayev University

Master of Public Health Program

"Conceptualization and Measurement of Food and Nutrition Literacy among University Students: A Systematic Review of Measurement Tools"

Master of Public Health Integrating Experience

Project Professional Publication Framework

By

Khassan Zhakay

MPH Candidate

Advisor/s: Kuanysh A. Yergaliyev, MD MPH DrPH, Assistant Professor of Practice

April, 2025

Abstract

Objective: Food literacy (FL) and nutrition literacy (NL) are both essential constructs in making informed dietary choices and navigating in complex food environments. Building these literacies among university students is crucial as they are in the pivotal stage of life when they have increasing autonomy over their food choices. The objectives of this systematic review were to (i) identify existing tools that measure FL and NL among university students and (ii) evaluate their psychometric properties using the COSMIN framework. **Design:** A systematic review was conducted following PRISMA guidelines. Sixteen studies were included based on eligibility criteria related to scale development and psychometric validation in university student populations and excluding studies that focus solely on dietary outcomes without discussing the underlying conceptualization or measurement of FL or NL. General and psychometric characteristics of tools were extracted and synthesized narratively. The COSMIN risk-of-bias checklist was applied to assess the psychometric rigor of the tools. **Setting:** Five academic databases were systematically searched in February 2025. **Participants:** All included studies involved university students, typically aged 18–25 years, from diverse geographical regions, including Asia, Europe, North America, and the Middle East. **Results:** Sixteen tools were identified, where 8 of them focused on NL, 6 on FL, 1 measured a critical component of NL, and 1 covered both. The reviewed tools differed substantially in terms of conceptual frameworks, item development methods, and psychometric properties. Most tools reported strong content, structural validities, and internal consistency; fewer studies assessed reliability, criterion validity, and hypothesis testing validity, and no tools evaluated cross-cultural validity and responsiveness. **Conclusion:** The systematic review demonstrates that recent FL and NL tools for university students have shifted to more contextually driven domains, such as critical thinking, sustainability, and digital literacy. While a majority of tools exhibit acceptable to strong

psychometric properties, gaps remain in the current theoretical and practical landscape of measurement tools being fragmented with inconsistent conceptual coverage of FL and NL and cross-cultural validation, indicating the need for further development of both conceptually and contextually appropriate tools.

Keywords: food literacy (FL), nutrition literacy (NL), health literacy (HL), university students, systematic review, measurement tools, COSMIN framework, psychometric validation

Acknowledgements

I would like to express my gratitude to my family and faculty for support and guidance through my master's degree with all its ups and downs.

Table of Contents

Abstract ii

Acknowledgements iv

List of Tables vii

List of Figures viii

Abbreviations ix

CHAPTER I: INTRODUCTION 1

1.1 Overview 1

1.2 Background of the study 2

1.3 Problem statement 3

1.4 Research aim and questions 4

1.5 Significance of the study 5

1.6 Outline of the study 5

1.7 Conclusion 6

CHAPTER II: LITERATURE REVIEW 8

2.1 Introduction 8

2.2 The Emergence of Food and Nutrition Literacy & Defining the Concepts 8

2.3 Measuring Food and Nutrition Literacy 14

2.4 The Quality Appraisal of Measurement Tools 18

2.5 Conclusion 21

CHAPTER III: METHODS 23

3.1 Introduction 23

3.2 Systematic review Framework 24

3.3 Eligibility Criteria 24

3.4 Information Sources and Search Strategy 25

3.5 Study Selection Process 25

3.6 Data Extraction 26

3.7 Quality Appraisal/Risk of Bias Assessment 26

3.8 Data Synthesis 27

3.9 Ethical Approval 27

3.10 Methodological Strengths and Limitations 28

3.11 Conclusion 29

CHAPTER IV: RESULTS 30

4.1 Conceptual frameworks 32

4.2 Item structure and used constructs 34

4.3 Development and validation methods 36

4.4 Sample characteristics 52

4.5 Cross-cultural adaptation 54

4.6 Quality appraisal of tools 73

4.7 Conclusion 84

CHAPTER V: DISCUSSION 85

CHAPTER VI: CONCLUSION 92

References 93

Appendices 107

Appendix A 107

Appendix B 109

Appendix C 110

List of Tables

Table 1: A summary table of FL and NL tools for university students..... 39

Table 2: The psychometric properties of FL and NL tools for university students..... 55

List of Figures

Figure 1. Food literacy definitions and conceptual frameworks by Krause (2016)..... 9

Figure 2. Nutrition Literacy definitions by Krause (2016)..... 10

Figure 3. Conceptual Hierarchy of Health, Food, and Nutrition Literacy..... 12

Figure 4. Food Literacy framework by Vidgen & Gallegos (2014)..... 13

Figure 5. PRISMA flowchart..... 32

Abbreviations

NL – Nutrition Literacy

FNL – Functional Nutrition Literacy

INL – Interactive Nutrition Literacy

CNL – Critical Nutrition Literacy

FL – Food Literacy

HL – Health Literacy

FHL – Functional Health Literacy

IHL – Interactive Health Literacy

CHL – Critical Health Literacy

UPF – Ultra-processed Foods

NVS – Newest Vital Signs

PCA – Principal Component Analysis

WHO – World Health Organization

BMI – Body Mass Index

EFA – Exploratory Factor Analysis

CFA – Confirmatory Factor Analysis

CVI – Content Validity Index

CVR – Content Validity Ratio

LPA – Latent Profile Analysis

CHAPTER I: INTRODUCTION

1.1 Overview

University students represent a unique demographic for the study of food and nutrition literacy as they are in the middle of transitioning into adulthood. At this time young adults experience substantial life-changing events such as leaving parental homes, moving away to university, and gaining autonomy in some aspects of life. It was noted that short- and longer-term dietary habits and lifestyle behaviors are formed at this stage of life (Papadaki et al., 2007) and there are new factors that can affect their dietary behavior such as increased "autonomy over food choice", "tight budgets", "exposure to new social groups", and "food cultures" (Sprake et al., 2018). Studies have shown that college students tend to acquire unhealthy habits when it comes to food choice. Their diet can be characterized by "low intake of fruit and vegetables, and a high consumption of energy-dense foods such as takeaway foods and sugar-sweetened beverages" (Ramírez-Contreras et al., 2021; Whatnall et al., 2019). All these factors in concrete with busy academic and social life may hinder acquiring healthy dietary behaviors and food skills. The consequences of an improper diet can increase the risk of obesity and associated comorbidities including diabetes and heart diseases (Alkazemi, 2019). Given the context of the university setting, one of the solutions for lowering the risk of obesity and adverse health outcomes can be food and nutrition literacy entailing educational interventions and initiatives. Unlike adults who have more established dietary habits and children who rely on their parental guidance for food choice, university students who are at a pivotal stage of acquiring food knowledge and dietary habits should be a main target for food and nutrition literacy-based health promotion and educational interventions. Therefore, the conceptualization and adequate measurement of these types of literacy among university students is of great importance for public health policies and initiatives.

1.2 Background of the study

Given the "complex food systems and technological advancements in the production of food, transportation methods, progress in nutrition research the ability to maintain health and well-being through food and nutrition has paradoxically become increasingly difficult" (Colatruglio & Slater, 2014). In light of these issues, the scientific community has strived to frame the concepts that can explain and measure complex systems of food and nutrition, giving an opportunity to effectively navigate in the field. One of the well-defined and famous concepts is food literacy (FL) which has been described by Vidgen & Gallegos (2014) as "the scaffolding that empowers individuals, households, communities or nations to protect diet quality through change and strengthen dietary resilience over time. It is composed of a collection of inter-related knowledge, skills, and behaviors required to plan, manage, select, prepare and eat food to meet needs and determine intake". According to prior studies, adequate food literacy (FL) and health promotion literacy are key factors in shaping young adult's diet (Rosas et al., 2020). Food literacy-based curriculum improved food literacy-based behaviors, self-efficacy and confidence among students at Appalachian State University (Morgan et al., 2023). The mobile Health food literacy intervention in Ontario, Canada had a positive impact on student's perception on healthy dietary behaviors (Schaafsma et al., 2025). According to (Kabasakal-Cetin et al., 2024), greater food literacy along with sustainable and healthy eating behaviors predicted higher unprocessed food consumption in contrast to lower food literacy along with sustainable and healthy eating behaviors which predicted higher ultra-processed foods (UPF) consumption among students aged 18-30 years.

In parallel, the nutrition literacy (NL) concept in relation to health literacy was developed. The NL is a "degree to which individuals can obtain, process, and understand the basic health (nutrition) information and services they need to make appropriate health (nutrition) decisions, with the qualification that the definition is nutrition-specific." (Silk et

al., 2008). Similar to FL, high NL predicted healthy eating behaviors among college students (Lai et al., 2021), nursing students (Mostafazadeh et al., 2024), students who have received health education (Iskender et al., 2024). NL, like FL, had a positive effect on a healthy diet among university students. It can be assumed that increasing levels of NL together with FL has a good practical application by integrating both into the university curriculum.

Both concepts have multiple definitions given by researchers in the field of nutrition. There is no consensus on only one right definition of FL and NL. But the majority of definitions point to the statement that NL is mainly about understanding nutrition information and its impact on health while FL is way broader in a way that it includes also acting on that information to facilitate healthy eating (Silva et al., 2023). It can also be supported by the following relationship between FL and NL where FL as a broader concept "is an ability to make decisions that lead to better individual health status and lead to a sustainable food system considering all social, environmental, cultural, economic, and political factors" (Cullen et al., 2015). Nutrition literacy is more a "subfield of food literacy", with both being dimensions of health literacy (Krause, Sommerhalder, Beer-Borst, et al., 2016). FL and NL are different but complementary concepts (Cullen et al., 2015; Krause, Sommerhalder, Beer-Borst, et al., 2016). Together, these concepts point to a holistic framework that emphasizes the importance of FL and NL for university students who are susceptible to poor dietary habits in acquiring not only factual knowledge, but also critical thinking, practical food skills, and an understanding of the larger food system.

1.3 Problem statement

The development and measurement of FL and NL is widely conducted and systematically reviewed among adults (Yuen et al., 2018), children and adolescents (Carroll et al., 2022). A mini-review of food literacy measurements among youth aged 16-24 years for

the previous 5 years was conducted by Groufh-jacobsen & Medin (2023), which included both high school and university student populations. However, to our knowledge, a systematic review specifically examining the conceptualization and measurement of both FL and NL within a university setting has not yet been conducted. Therefore, the target of the systematic review will be specifically individuals with status as students. Considering the importance of students as an intervention demographic to promote healthy dietary behavior and protect them from acquiring unhealthy dietary patterns, a systematic review that consolidates existing tools and critically examines them in accordance with conceptual frameworks is both timely and warranted in the university context. Therefore, the aim of the study is to conduct a systematic review of studies conceptualizing FL and NL among university students and their related measurement tools.

1.4 Research aim and questions

Therefore, the aim was to conduct a systematic review to examine how food literacy (FL) and nutrition literacy (NL) are conceptualized and measured among university students, including an evaluation of the quality and psychometric properties of the measurement tools used.

Research Questions:

- Primary Research Question: How are food literacy (FL) and nutrition literacy (NL) conceptualized and measured among university students?
- Secondary Research Question: What are the psychometric properties (validity, reliability) of the measurement instruments used to assess FL and NL among university students, as evaluated using COSMIN risk of bias checklist?

1.5 Significance of the study

The systematic review of current conceptual frameworks of FL and NL, and measurements based on them among university students can contribute to the increasing body of academic knowledge by elucidating theoretical and methodological approaches used in the nutrition field to measure food and nutrition literacy. By synthesizing existing literature, this study strives to clarify conceptual frameworks mainly used in university student samples to assess their level of FL and NL, enhancing theoretical clarity and aiding future research. The review of psychometric values of existing tools and their quality appraisal may identify gaps in reliability, validity, and applicability, to refine existing tools or to develop new standardized, cross-cultural tools. The findings can have implications for designing FL, NL-based interventions in universities, as improving FL and NL has been linked to healthier dietary choices and reduced risk of obesity and related comorbidities.

1.6 Outline of the study

The master's research thesis consists of six main chapters, which are: Introduction, Literature review, Methodology, Results, Discussion and Conclusion. In Chapter 1, the problem statement articulates the purpose of the study and outlines the research aim and questions. It also underlines the relevance of Food Literacy (FL) and Nutrition Literacy (NL) within the context of higher education and discusses the potential implications of the study for educational program design and health promotion strategies among university students. Chapter 2 begins with an exploration of theoretical definitions and conceptual frameworks underlying FL and NL, followed by an analysis of the role of these literacies in shaping dietary behaviors of university students. Focus is given to the existing measurement tools developed to assess FL and NL, and due to absence of comprehensive systematic reviews on measurement of these literacies among university students, the literature review of the

existing systematic reviews on other age groups were conducted, including their content domains and psychometric properties. The chapter concludes with the assessment of current challenges and gaps in the measurement of FL and NL in university settings. Chapter 3 outlines the methodological approach adopted in the study. A systematic review design was employed with the primary aim of identifying, appraising, and synthesizing measurement tools used to assess FL and NL among university students. The search strategy, eligibility criteria, data extraction process, and the use of the modified COSMIN risk of bias checklist for psychometric evaluation are described in detail (Mokkink et al., 2018). All ethical considerations related to systematic data handling were addressed in accordance with institutional review standards and due to usage of secondary, non-personal data, the institutional review approval was not obtained. Chapter 4 presents the findings of the systematic review including a synthesis of the general characteristics, conceptual frameworks of the identified FL/NL tools and their psychometric properties. The findings are organized into tables summarizing (1) tool general characteristics and (2) psychometric properties. Each tool is evaluated according to the modified COSMIN risk of bias checklist criteria with strengths and limitations systematically appraised. In the following Chapter 5 the study findings that emerged during the systematic review process were discussed in relation to the literature that was relevant to the purpose of the study. In the final chapter 6, there were provided a summary of the key findings addressing the research questions, acknowledgement of the limitations of this study, and policy and practice implications for the main stakeholders, commencing by laying out suggestions for further research.

1.7 Conclusion

To sum up, this chapter has discussed the background and significance of food and nutrition literacy within the context of higher education. The goal of the study is to

systematically review and critically evaluate existing FL and NL measurement tools in terms of their conceptual foundations, development processes, and psychometric properties. Addressing this research gap becomes imperative, because university students represent a pivotal population for long-term dietary habit formation and the current theoretical and practical landscape of measurement tools remain fragmented with inconsistent conceptual coverage of FL and NL, and limited validation. Therefore, the following chapters will elaborate on the further elements of the study. Literature review revealed that university students are a vulnerable group with a high risk of acquiring unhealthy dietary habits during their studies. On this occasion defining the existing concepts of FL and NL as well as their adequate measurements among university students is crucial. There is a limited amount of knowledge on FL and NL and their measurement among university students compared to adults, children, and adolescents. That is why a systematic review of FL and NL concepts and measurements among university students is needed.

CHAPTER II: LITERATURE REVIEW

2.1 Introduction

Prior to conducting this study, a review of the literature on FL and NL concepts and measurement scales was carried out using various research databases—such as PubMed, Web of Science, ProQuest, CINAHL, and MEDLINE—and included peer-reviewed articles, electronic books, and grey literature sources (e.g., theses and dissertations). The existing body of knowledge on topic indicated that high FL and NL were associated with improved dietary habits, positive perception of healthy dietary choices, and less consumption of ultra-processed foods (UPF) among university students (Iskender et al., 2024; Kabasakal-Cetin et al., 2024; Lai et al., 2021; Morgan et al., 2023; Mostafazadeh et al., 2024; Schaafsma et al., 2025). Interventions based on FL and NL concepts are crucial during university time as they can prevent young adults from acquiring unhealthy dietary patterns.

The definitions and conceptual frameworks of FL and NL are frequently cited in concert with HL model by Nutbeam (2000) and FL framework of Vidgen & Gallegos (2014); and there are numerous measurement tools differing from each other by target population, theoretical frameworks, psychometric values, and cultural characteristics. The literature review will include the current state of knowledge on that and try to explore the research landscape on FL and NL conceptualization and measurements.

2.2 The Emergence of Food and Nutrition Literacy & Defining the Concepts

The term “literacy” has traditionally been linked to basic reading and writing skills, but its application in the health domain has evolved to include the ability to understand, interpret, and apply health-related information (Nutbeam, 2000). Within this broader framework, FL and NL have emerged as independent but overlapping concepts (Velardo, 2015). FL has been characterized as the ability to “plan”, “select”, “prepare”, and “eat”

healthy food, whereas NL focuses more particularly on knowledge about nutrients, dietary requirements, and the health implications of food choices (Vettori et al., 2019; Vidgen & Gallegos, 2014). It was recommended by Velardo (2015) to use already established and closely related term of health literacy (HL) by Nutbeam (2000) in defining FL and NL. According to Nutbeam's HL framework, there are three types of literacy: functional, interactive, and critical. Functional HL (FHL) is an ability to obtain, understand, and use factual health information. Interactive HL (IHL) is an ability to communicate effectively to obtain, provide and apply health information to improve health. Critical HL (CHL) includes the ability to critically assess and reflect on health information in their social context (Nutbeam, 2000). Krause, Sommerhalder, Beer-Borst, et al., (2016) systematically reviewed existing theoretical definitions of FL and NL against three HL types.

Food Literacy Definitions	Author, Year	Literacy Component		
		Functional	Interactive	Critical
Conceptual frameworks of Food Literacy	Fordyce Voorham, 2011	X		
	Kolasa et al., 2001	X		
	Eat well South Australia, 2010	X		
	Thomas and Irwin, 2011	X		
	Slater, 2013	X	X	X
	Sustain, 2013	X		X
	Schnoegl et al. 2006	X		X
	Vidgen and Gallegos, 2014	X	X	X
	Howard and Brichta, 2013	X		X
	Topley, 2013	X	X	X
	Desjardins and Azevedo, 2013	X	X	X
	Smith, 2009a	X	X	X
	Block et al. 2011	X		X

Figure 1. Food literacy definitions and conceptual frameworks by Krause (2016)

Seven conceptual frameworks and six definitions of FL were described as a result of a systematic review conducted by Krause et al. (2016) (Fig 1)—all conceptual frameworks described both functional and critical dimensions of HL. Functional FL involves knowledge and range of skills starting from basic cooking skills, planning and budgeting, understanding of the effect of food choice on health and well-being (Krause, Sommerhalder, Beer-Borst, et

al., 2016). Interactive FL included food literacy definition of Slater, (2013) which is based on the assumption that food and nutrition knowledge builds personal skills such as decision-making and goal-setting. Conceptual framework by Smith (2009) includes several interactive elements such as "sharing life experience; empathizing with others ('lifeworld food literacy'); cooperative learning ('interactive/interpretive food literacy'); and, using storytelling and narratives to explore the meanings of food ('narrative food literacy')." Vidgen & Gallegos, (2014) framework denoted "gathering and eating in a social way" and Desjardins & Azevedo, (2013) framework included the ability to "share information and transfer skills" as the part of the interactive dimension of FL.

Regarding critical FL competencies conceptual frameworks identified three areas: (1) judging the credibility of nutrition information; (2) reflecting on social, cultural, and situational factors that shape dietary behaviors and; (3) how food choice affects broader society (Krause, Sommerhalder, Beer-Borst, et al., 2016).

	Author, Year	Literacy Component		
		Functional	Interactive	Critical
Nutrition Literacy Definitions	Blitstein and Evans, 2006	X		
	Guttersrud et al., 2014	X	X	X
	Neuhauser et al., 2007	X		
	Silk et al., 2008	X		
	Watson et al., 2013	X		
	Zoellner et al., 2009	X		

Figure 2. Nutrition Literacy definitions by Krause (2016)

Six definitions and no conceptual frameworks of NL were identified after a systematic review. All NL definitions contained functional characteristics of HL except for definition by Guttersrud et al. (2014) where the NL encompassed all dimensions of HL (Fig 2). The definitions are mainly organized around "cognitive capacities" and "basic literacy and numeracy" skills (Krause, Sommerhalder, Beer-Borst, et al., 2016). The NL definitions offered by Neuhauser et al. (2007) and Watson et al. (2013) included the ability to interpret

food labeling on the front package and menu label, and understand basic nutrition information as functional nutrition literacy. Interactive and critical NL from Guttersrud et al., (2014) definition is stated as "the ability to evaluate the quality of nutrition information; and the willingness to take action to improve nutrition health in families, communities, or broader social and global movements."

The systematic review of Krause et al., (2016) revealed that there are definitions and conceptual frameworks of FL and only definitions of NL. The researchers state that there is a lack of interactive domains of FL frameworks. For future research, the coverage of all HL dimensions can help develop comprehensive measuring tools and conceptual frameworks that can be widely used in the scientific community. Because of the absence of consensus on conceptual frameworks and numerous definitions it is quite complicated to conduct educational interventions and studies as researchers are overwhelmed by the number of various definitions, concepts, and their interpretations. This conclusion is also supported by a similar systematic review by Truman et al. (2017) where food FL definitions were analyzed against six thematic areas: "Skills/Behaviors", "Food/Health choices", "Culture", "Knowledge", "Emotions", and "Food Systems". The difficulty lies in the diversity of evidence collected which makes it difficult to provide a definitive conception of "food literacy" (Truman et al., 2017). Another aspect that needs to be addressed is to explore the roles of attitudes, motivation and behaviors in FL. Behavioral changes were addressed in later study by Thomas et al. (2019) where FL includes self-efficacy and confidence as one of the components of "interconnected attributes organized into the categories of food and nutrition knowledge; food skills; food decisions; and ecologic (external) factors".

Both FL and NL coexisted in the literature and differences between them were unclear which led to the inability to measure the effects of FL and NL. Based on the results of the systematic review conducted by Krause et al. (2016), skills involved in NL are a prerequisite

for acquiring skills in FL domains but not sufficient to stand alone in navigating complex dietary and nutrition systems. All definitions of NL and half of FL were based on HL (Krause, Sommerhalder, Beer-Borst, et al., 2016). The FL and NL are both forms of HL rather than separate concepts and NL is a subfield of FL (Krause, Sommerhalder, Beer-Borst, et al., 2016). Based on literature and existing definitions, the scope of concepts can be depicted as follows:

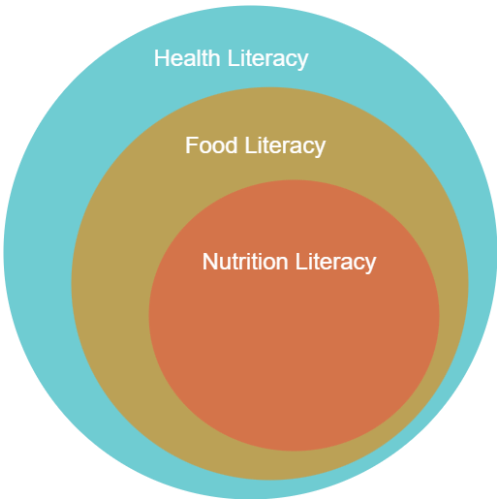

Figure 3. Conceptual Hierarchy of Health, Food, and Nutrition Literacy. *Note. Figure created by the author.*

The hierarchical relationship between HL, FL, and NL can illustrate how FL and NL are nested within broader literacy domains, reinforcing the idea that NL is a critical but insufficient component of FL.

Despite the absence of consensus and ever-growing amount of knowledge on FL, it is important to mention well-known FL framework by Vidgen & Gallegos (2014) that empirically defines and analyzes the term 'food literacy' for the first time as indicated by the authors and show robustness by covering all three levels of literacy. The framework identified eleven components of FL across four dimensions: Plan and Manage, Select, Prepare, and Eat (Fig 4). Sample included disadvantaged young adults and identified the importance of

improved FL on empowering various societal levels to enhance diet quality. The conceptual framework is one of the fundamentals in framing the FL and further advancements of research in the diet and nutrition field.

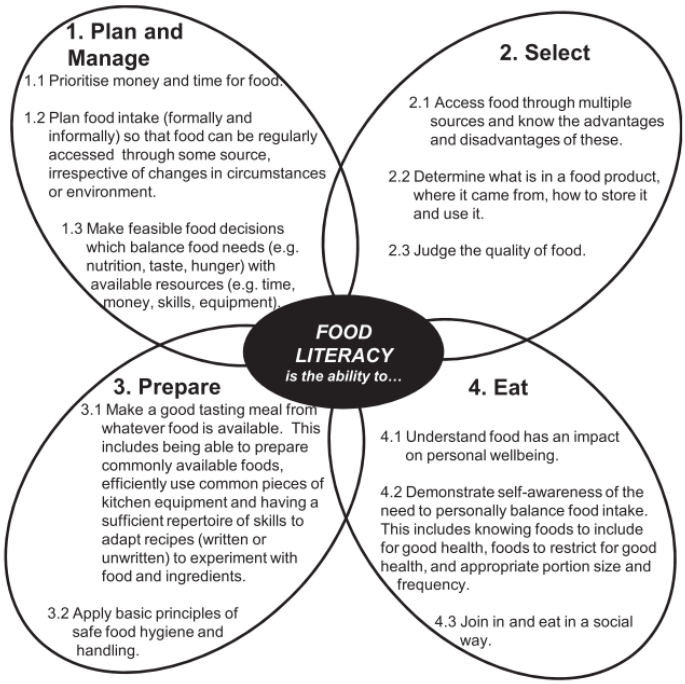

Figure 4. Food Literacy framework by Vidgen & Gallegos (2014)

Not so many variations in NL conceptualization, however way more conceptual frameworks along with definitions exist about FL. Recently, FL definitions were supplemented with 38 novel definitions by Truman et al. (2017), and 51 more novel definitions by Thompson et al. (2021). They are comprehensive and well-crafted; however, they can differ from each other based on the scope, where one definition has a broader while the other one has a narrower perspective on diet and nutrition, or the subconstructs within the definitions can vary greatly causing differences and inconsistencies. Considering that dietary and nutrition knowledge is still evolving, the consensus-based approach in developing comprehensive theoretical frameworks in conceptualizing FL and NL is important.

2.3 Measuring Food and Nutrition Literacy

The research field on NL assessment tools often references the foundational study by Weiss et al. (2005) on the Newest Vital Sign instrument (NVS) available in English and Spanish. The instrument aimed to quickly measure limited literacy in primary healthcare settings. The NVS like Nutrition Label Survey by Rothman et al. (2006) focused to measure the ability to use food labels. The NVS established an early model for future nutrition literacy measures that tried to achieve a balance between reliability and practicality of administration. Further advancements in measuring literacy were marked by the work of Diamond (2007), who measured NL among adults by developing the Nutrition Literacy Scale (NLS). NLS demonstrated an acceptable internal consistency and addressed the NL measurement among adults in response to growing nutrition-related conditions such as obesity and diabetes.

Recent studies on FL and NL measurements reflect a transition from just knowledge-specific tools to more comprehensive, context-specific instruments that include multiple dimensions of literacy. While earlier studies emphasized quick assessments (NVS) or basic nutrition knowledge (NLS), the newly emerging studies acknowledge that literacy extends beyond just functional reading and comprehension. Researchers increasingly recognize that being 'literate' in food and nutrition also involves interacting with information, applying it in everyday contexts, and critically evaluating a wide range of factors that shape dietary choices. One of the studies that embraced this approach was Guttersrud et al., (2014) who assessed CNLs (Critical Nutrition Literacy) among nursing students. The study explores the critical dimension of NL beyond functional. The CNLT (Critical Nutrition Literacy Tool) developed by Guttersrud et al. (2014) later revised by McNamara et al. (2020) and was used to measure CNL of 1700 U.S college students. The NL tools are not limited to them as there are different tools with different target populations and purposes such as NLit marked as psychometrically robust (Gibbs & Harvey, 2017) and its variations: NLit-BCa (Gibbs et al.,

2016), NLit-S (Gibbs, Camargo et al., 2018), NLit-P (Gibbs, Kennet et al., 2017). Electronic assessment tools such as e-NutLit (Ringland et al., 2016) and the Japanese NL tool NLQ-JP (Aihara & Minai, 2011).

FL measurement tools identified by Jo et al. (2021) mainly focus on specific populations or are adaptations of existing tools adjusted for cultural and contextual nuances. Instruments for children from Melbourne (Wijayaratne et al., 2022), Australian adults (Thompson et al., 2022), Tehrani school-children (Doustmohammadian et al., 2022), and Spanish university students (Luque et al., 2022). Culturally adapted ones such as the Polish version of the Short Food Literacy Questionnaire (SFLQ) (Zwierczyk et al., 2022), the Turkish version of SFLQ (Durmus et al., 2019), and the Brazilian version (Zeminian et al., 2022) provide measurement outcomes that can't apply to other than specific groups (Fisher & Potgieter, 2024). While this demonstrates versatility, it also indicates the absence of a single comprehensive, validated FL scale. The absence of a universal measuring scale is further complicated by the variety of FL frameworks and definitions. According to Fisher & Potgieter, (2024), the first step towards the development of a universal FL scale is to reach an agreement on the FL definition and its subconstructs which will be context and culture-specific.

The systematic review by Yuen et al. (2018) consolidating information about thirteen instruments of NL and FL, concluded that NLit had the strongest psychometric properties (Gibbs et al., 2017). The review highlighted the need for comprehensive NL and FL instruments to evaluate interventions effectively. According to Yuen et al. (2018) only three out of thirteen measures were based on working definitions of FL and NL, none of the tools assessed responsiveness over time, only three tools had tested for test-retest reliability, and only eight measures included scoring methods to differentiate between various levels of literacy which indicates low methodological rigor. Also in review, NL scales are predominant

compared to FL scales (Gréa Krause et al., 2018; Palumbo et al., 2017) among adults even though the FL construct is broader. On this matter, Yuen et al. (2018) suggested that more research are needed to develop psychometrically robust measures designed to assess FL and its key domains, including “volitional and behavioral factors”.

The scoping review of FL measurement tools among adults by Amouzandeh et al. (2019) identified twelve tools. Five of them were explicitly developed to measure the FL construct. Developed for adult populations in Italy (Palumbo et al., 2017), Switzerland (Gréa Krause et al., 2018), Australia (Begley et al., 2018), the US (Lahne et al., 2017), and the Netherlands (Poelman et al., 2018) these tools were largely guided by Vidgen & Gallegos (2014) framework except for Switzerland adults (Gréa Krause et al., 2018). Although the four of the tools shared a common conceptual basis, their individual items varied considerably. This variation reflected the complexity of measuring FL and reaching a consensus that would support comparative measurement (Fingland et al., 2021). Considering this complexity the adoption of a universally applicable FL tool should be one of the main goals in the field. Amouzandeh et al., (2019) stated the following factors as limitations of study: most tools capture the four domains of FL but fail to address all eleven components as conceptualized by Vidgen & Gallegos (2014), existing tools have been validated in limited contexts and demographics and often rely on self-report methods increasing the risk of bias. The construct validity of some FL tools was tested against factors such as education and income, though these variables may not fully grasp the construct since people with different socio-economic backgrounds can be food literate. Moreover, the tool’s content validity was tested only by professionals – dietitians, and public health experts – without involving the general population, despite food literacy being used in everyday practice. Additionally, the validation samples were typically homogeneous, consisting mainly of highly educated females from Western countries, limiting the tools' applicability across diverse contexts and making cross-

group comparisons difficult. Although no single tool stands out as the best, the review highlighted that using multidimensional, contextually validated FL measurement tools is important. Tools that capture the full range of FL dimensions and components—such as those developed by Begley et al., Lahne et al., Palumbo et al., Poelman, and Wijayaratne et al.—are better suited to reflect the construct's interrelated nature (Amouzandeh et al., 2019). Validity is further enhanced when instruments are selected based on their relevance to specific populations and context, yet current validations are limited to general adult populations, low socio-economic groups, household food gatekeepers, and adults with cancer in Western countries. Significant gaps exist, particularly in terms of integrating these methods across varied contexts and including performance indicators at the population level—an important concern given that FL spans not only individuals but also societies.

The FL and NL measurements among children and adolescents were systematically reviewed by Carroll et al. (2022). According to the results, twelve tools were identified, where three of them measured FL, four measured both, and the remaining, NL and its subdomains, specifically, CNL, menu board literacy, and food label literacy (Carroll et al., 2022). Authors highlighted that FL and NL are highly contextual constructs based on the food's system geographical, cultural, and social features. That's why measurement tools are also highly contextual. Also, the FL and NL definitions do not address the age-specific criteria that are mostly ignored by FL and NL definitions. The Plan and Manage domain of Vidgen & Gallegos, (2014) framework does not apply to children and adolescents with limited independence over food choices. "Preschool-FLAT" by Tabacchi et al. (2020) stated to have modifications on Vidgen & Gallegos' (2014) framework to cover knowledge and skills suitable for preschoolers but provided limited details on how it was done. Other tools also did not incorporate the FL dimensions suitable for youth developed by Slater et al. (2018).

The development of FL and NL measures underwent a transition from narrowly focused, functional knowledge-based tool development to more comprehensive, beyond functional, context-based approaches in crafting measurement instruments. Early instruments like NVS and NLS aimed to quickly assess basic literacy skills, especially food label reading and comprehension abilities. Advancements in this field led to a great variety of FL and NL tools based on geographical location, target population, health conditions, age groups, social status, etc. All of them covered a broader range of nutrition knowledge and skills. However, despite these advancements, the field of nutrition continues to face challenges, such as the lack of universally applicable, psychometrically robust FL and NL scales, inconsistencies in theoretical frameworks and definitions, and limited validation across diverse populations and age groups. Addressing these gaps is crucial for developing assessment instruments that can be effectively utilized in educational interventions and public health initiatives.

2.4 The Quality Appraisal of Measurement Tools

The critical appraisal of FL and NL scales is important for ensuring their validity, reliability, and appropriateness in various contexts. The process of critically appraising involves evaluating validity, reliability, responsiveness and statistical rigor of these tools to ensure they accurately measure the intended constructs (Mokkink et al., 2018). Overall, two systematic reviews, one on children and adolescents, and one on adults were conducted yielding approximately 20-25 measuring tools (Carroll et al., 2022; Yuen et al., 2018). A scoping review by Amouzandeh et al. (2019) described twelve more tools for adults.

The list of tools mentioned in Yuen et al., (2018): NLS, Spanish NLS, NVS, NLit, NLit-BCa, NLit-S, NLit-P, CNLI, e-NutLit, SFLQ, NLQ-JP, CHLSalt-HK, IT-FLS. Only two of them were based on underlying conceptual framework (SFLQ and IT-FLS). SFLQ is based on FL and NL components measurement against HL dimensions (Krause, Sommerhalder,

Beer-Borst, et al., 2016) while IT-FLS is based on (Vidgen & Gallegos, 2014) framework. Majority of tools demonstrated good content validity except for NLS and NLQ-JP where intended domains included not clearly stated. The content validity is the degree to which the content of an instrument is an adequate reflection of construct described (Mokkink et al., 2014). Also, NLS had partial inclusion of domains because it focused only on functional literacy such as reading comprehension. SFLQ failed to cover all twelve components, e-NutLit, and NLQ-JP had unclear descriptions about components inside domains of FL and NL. Majority of tools had good face validity. The construct validity measures do the instrument measure in the same way as other instruments/tools/indices. The construct validity analysis showed that CNLI, SFLQ, NLQ-JP, CHLSalt-HK did not compare with other NL/HL/literacy indices. Other tools had a partial positive relationship/correlation with other tools such as S-TOFHLA (Short Test of Functional HL in Adults) for basic reading, writing, numeracy in healthcare, HEI-2010 (Healthy Eating Index-2010), SAHL-S (Short Assessment of Health Literacy-Spanish) (S. Y. D. Lee et al., 2010). Reliability is the extent to which the outcome measures are the same for repeated measurements. There is test-retest reliability which assesses whether the tool produces the same results with the same participants under the same conditions over a short period of time (Mokkink et al., 2014). Test-retest reliability of only three tools was measured: NLit, the test reliability during retesting ranged from "0.51 to 0.80 in six areas; the overall scale of test reliability during retesting was adequate – 0.88"; "NLit-BCa, the test-retest reliability ranged between fair and substantial across six domains (0.44-0.90)"; and CHLSalt-HK test-retest reliability was "adequate (n=41) (ICC (Item Characteristic Curve) = 0.85; 95% CI [0.707,0.919])". None of the tools were tested for inter- and -intra-rater reliability. Internal consistency across tools ranged from 0.76-0.97.

The FL and NL tools for children and adolescents were tested for validity and reliability. Five tools were validated for most types of validity, including content, face and

construct FNLIT (Doustmohammadian et al., 2018), FLI (Stjernqvist et al., 2021), M-FNLIT (Khorramrouz et al., 2022), FNLAT (Ashoori et al., 2020), Thai-NLAT (Deesamer et al., 2020) showing a methodologically rigorous process in designing these tools (Carroll et al., 2022). Majority of tools demonstrated acceptable to good internal consistency with a range of 0.77-0.90. However, six tools FNLIT (Doustmohammadian et al., 2018), FLI (Stjernqvist et al., 2021), TFLAC (Amin et al., 2019), FNLQ-SC (T. Liu et al., 2021), FNLAT (Ashoori et al., 2020), Thai-NLAT (Deesamer et al., 2020) had subscales with slightly poor and doubtful internal consistency scores. It was suggested by researchers that further adaptations may be needed to improve internal consistency among these subscales. Six out of twelve tools assessed test–rest reliability: FNLIT, FLI, TFLAC, M-FNLIT, FNLAT, FFLANK (Reynolds et al., 2012) with ICC (Intraclass Correlation Coefficients) ranging from 0.64 to 0.93 indicating moderate to excellent reliability. The Thai-NLAT (Deesamer et al., 2020), the FNLAT (Ashoori et al., 2020), the FNLIT (Doustmohammadian et al., 2018) and its adaptions M-FNLIT (Khorramrouz et al., 2022) had the strongest psychometric properties. However, factors such as the context and culture during development of these tools should be considered as the FL and NL are highly contextual, influenced by the geography and its socio-cultural context.

The quality appraisal of FL and NL measurements among adults can be complemented by scoping review of FL measuring tools by Amouzandeh et al. (2019). Content validity was measured in tools by Barbour et al. (2016), Begley et al. (2018), Boucher et al. (2017), Palumbo et al. (2017), Poelman et al. (2018), Wallace et al. (2016) by dieticians and public health specialists. The internal consistency values range from 0.76-0.95, which shows excellent consistency. The highest internal consistency was observed in scale by Wallace et al. (2016) for the confidence, in “cooking”, “shopping”, “planning” and “purchasing”. Cronbach’s $\alpha = 0.912$ for the general food literacy scale in the Palumbo et al.,

(2017) tool. The lowest internal consistency was observed for “selection”, “plan and manage”, and “preparation” scales in the tool by Begley et al. (2018). Amouzandeh et al., (2019) concluded the systematic review by providing five tools which demonstrated excellent content, face and construct validity, as well as reliability (Begley et al., 2018; Gréa Krause et al., 2018; Palumbo et al., 2017;Lahne et al., 2017; Poelman et al., 2018)

2.5 Conclusion

The literature review has explored various factors and challenges that influence the conceptualization and assessment of Food and Nutrition Literacy (FL and NL) among university students, including theoretical inconsistencies, variability in measurement tools, and limitations in cross-cultural applicability. In conclusion, the overviewed literature suggests that the development and validation of FL and NL assessment instruments require further refinement, as university students represent a critical demographic for health promotion initiatives. Moreover, additional research is needed to provide evidence-based recommendations for enhancing literacy-focused educational interventions and integrating these constructs into university curricula.

The literature review including the quality appraisal of existing and widely used tools was conducted. Available studies have shown that the measuring tools were reviewed rigorously in various age groups: children and adolescents, adults; different socio-economic groups, low-income adults, among household food gatekeepers; and different health status such as adults with cancer in Western countries, etc. However, there remains a gap in consolidating the measurement tools for food literacy (FL) and nutrition literacy (NL), as well as in appraising their quality among university student populations. Based on the literature review of existing definitions and conceptual frameworks of FL and NL, and the tools that have been crafted to measure them, it is notable that there are also tools for

university students among them (Luque et al., 2022; McNamara et al., 2022) and they are poorly reviewed in relation to other segments of population.

There is an increasing body of knowledge on FL and NL and their measurement which is reflected in a variety of definitions and conceptual frameworks. There is no consensus on a comprehensive and overarching definition or conceptual framework when it comes to FL and NL. However, there are widely cited works by Vidgen & Gallegos (2014) on the conceptual framework of FL, and Nutbeam's (2000) tripartite health model which is considered a fundamental concept from FL and NL were derived. There is a great variety of definitions and frameworks with a different scope and components. The literature indicates the importance of universal and comprehensive definitions and frameworks for future educational interventions and public health initiatives.

The quality appraisal of measurement tools is an important step towards developing and validating comprehensive and universal tools. The literature review on FL and NL scales among children and adolescents, adults, and youth showed moderate to good scores in validity and reliability measures but still lacks validation in different demographics and inapplicability of culture, country-specific tools on wider population.

CHAPTER III: METHODS

3.1 Introduction

The primary goal of this study is to systematically review and critically evaluate the conceptual frameworks and measurement tools of Food Literacy (FL) and Nutrition Literacy (NL) developed for university students, and to answer the following research questions:

1. How are food literacy (FL) and nutrition literacy (NL) conceptualized and measured among university students?
2. What are the psychometric properties (validity, reliability) of the measurement instruments used to assess FL and NL among university students, as evaluated using COSMIN risk of bias checklist?

To address these research questions, a systematic review methodology based on PRISMA was conducted. A systematic review is a method that provides the comprehensive, transparent, and replicable synthesis of existing evidence on a clearly formulated topic using pre-defined inclusion/exclusion criteria. This method was selected for several reasons. First, the field of FL and NL measurement is characterized by a growing yet fragmented body of literature with a diversity of tools developed in different contexts and guided by heterogeneous conceptual foundations. A systematic review is a convenient method for consolidating this body of knowledge giving an opportunity for a summarized comparison of measurement tools and identification of methodological strengths and limitations. Second, a systematic review aligns well with practices in psychometric research where quality appraisal of measurement properties is important for guiding future tool development. Other methodologies are hard to implement considering the heterogenity of measurement tools and their characteristics. Employing a standardized evaluation framework, such as the modified COSMIN risk of bias checklist (Mokkink et al., 2018) ensures the rigor during quality appraisal process. Finally, the method provides a basis for drawing conclusions about the

current state of FL and NL assessment in higher education settings, informing researchers, educators, and policymakers interested in health promotion and educational intervention aimed at increasing FL and NL among university students populations. Additionally, the approximate frequency of usage of reviewed tools was measured to identify the widely employed tools in the field to support the quality appraisal section. The search was conducted in the same academic databases using the names of tools. The relative utility of the tools is summarized in supplementary table 2 in Appendices part of the thesis project.

3.2 Systematic review Framework

The systematic review is conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) to explore how Food Literacy (FL) and Nutrition Literacy (NL) are conceptualized and measured among university students. The goal of the systematic review is to map existing instruments and frameworks on which they were based, examine their psychometric properties using the COSMIN risk of bias checklist (Mokkink et al., 2018), and identify gaps in current measurement practices.

3.3 Eligibility Criteria

Eligibility criteria for the systematic review were defined using the SPIDER framework (Cooke et al., 2012), tailored to the conceptual and measurement focus of FL and NL within university populations. Studies were considered eligible if they involved university or college students (generally aged 18–25) or if data specific to this demographic or age range that can be extracted as sample. The studies had to explicitly define, conceptualize, or measure FL and/or NL, reporting on the development, validation, or application of relevant measurement instruments as phenomenon of interest. The systematic review included

empirical studies reporting on instrument development, validation, or conceptual analysis that presented psychometric data on the reliability, validity, or responsiveness of tools as design. Psychometric properties and theoretical robustness of the measurement tools as evaluation and quantitative validation or mixed-methods studies as research types. Studies focusing only on dietary outcomes or behavioral changes without paying attention to how FL or NL were measured were excluded, as were those involving populations outside the university context unless student-specific results were clearly defined. Only English-language publications were considered, with no strict time limit, however, most included studies were expected to be published post-2000 due to the relatively recent emergence of FL and NL concepts.

3.4 Information Sources and Search Strategy

A systematic search was conducted in multiple databases including PubMed, MEDLINE (via EBSCOhost), ProQuest, Web of Science, and CINAHL (via EBSCOhost). The search strategy involved search codes with keywords ‘food literacy’, ‘nutrition literacy’, ‘student*’, ‘university student*’, ‘tool*’, ‘scale*’, ‘instrument*’ to capture the breadth of FL and NL measurement. Terms were combined using Boolean operators like AND, OR and adapted to each database’s indexing system. To reinforce the search, the reference lists of included articles were also examined for additional related studies. Grey literature sources, such as dissertations and theses, conference papers were also considered if they presented new psychometric evidence on measurement data relevant to FL or NL among university students. All the data on database search and search codes are presented in Appendices.

3.5 Study Selection Process

The study selection process consisted of two sequential screening levels. First, two independent reviewers (K.Z. and K.Y.) screened the titles and abstracts retrieved from the

database search. At this initial screening stage, any study that clearly did not meet the inclusion criteria—due to irrelevance, wrong target population, no data on measurement of FL and NL, only dietary outcomes measure after FL and NL interventions were excluded. Potentially eligible records were then advanced to a full-text review, during which the same two reviewers independently assessed the articles against the prespecified inclusion/exclusion criteria. Disagreements or uncertainties concerning inclusion were resolved through discussion or, when necessary, by consulting a third reviewer.

3.6 Data Extraction

The data extraction procedure included collecting information covering bibliographic details such as authors, publication year, and country; key study characteristics such as design, sample size, demographic information of participants. Emphasis was placed on how FL or NL was defined, including any theoretical frameworks underpinning those definitions. Information on the measurement tools themselves—name of the scale, target domains or constructs, item count—was systematically recorded, along with any reported psychometric properties such as reliability (Cronbach's alpha, test-retest coefficients), and validity (content, construct, criterion, cross-cultural).

3.7 Quality Appraisal/Risk of Bias Assessment

The risk of bias was evaluated using the COSMIN risk of bias checklist (Mokkink et al., 2018). The assessment was conducted by two reviewers (K.Z and K.Y) independently. The COSMIN checklist consists of ten areas of assessment: patient-related outcome measure development, content validity, structural/construct validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness. To assess the methodological quality of each study, the

corresponding COSMIN risk of bias box (area) should be completed. Each area is evaluated with ‘Very Good’, ‘Adequate’, ‘Doubtful’, ‘Inadequate’, ‘Not Applicable’, and ‘Not reported’. To determine the overall quality of a study, the lowest rating of any area is taken based on the “worst score counts”. For example, if for an internal consistency study, one item in a box is rated as ‘inadequate’, the overall methodological quality of that internal consistency study is rated as ‘inadequate’. In the case of ‘Not Applicable’ and ‘Not reported’, the worst score count is not applied. The patient-related outcome measure development and content validity boxes were not included in this systematic review because these two boxes are based on criteria to assess clinical outcome measures not applicable to cognitive-behavioral assessment of literacy measures in university settings.

3.8 Data Synthesis

Considering the possible heterogeneity of data on the conceptualization of FL and NL and in the study designs and types of instruments used, a narrative synthesis approach was conducted. First, descriptive tables were generated to summarize each included study’s key features such as authors, samples, instrument details, and psychometric findings. These tables provided a comparative view of measurement approaches, enabling the identification of common themes, gaps, and inconsistencies across literature.

3.9 Ethical Approval

Ethical approval was not required for this review since it solely involved secondary analysis of already published data.

3.10 Methodological Strengths and Limitations

The methodological strengths of adopting systematic review are comprehensive and replicable steps to summarize and evaluate FL and NL measurement tools. First, using a predefined systematic review protocol according to PRISMA (Page et al., 2021) can minimize selection bias and enhance the methodological rigor during evidence synthesis. Second, the adoption of the COSMIN risk of bias checklist (Mokkink et al., 2018) along with PRISMA to assess the psychometric properties of reviewed scales can increase the objectiveness and reproducibility of the evaluation processes contributing to the methodological rigor of the study. Third, the research question and aims based on the SPIDER framework have provided solid inclusion/exclusion criteria for systematic review. Accordingly, studies corresponding to defined sample, phenomenon of interest, design, evaluation, and research types were included for systematic review. Along with methodological strengths, certain limitations were raised during the systematic review, such as the review being restricted to peer-reviewed articles published only in English, introducing language bias. While efforts were made to ensure that all eligible studies were included during the database search, the reliance on five academic databases may have limited completeness of retrieval, particularly when it comes to grey literature. The heterogeneity of the included studies in terms of conceptual coverage, item numbers and structures used in scales, and psychometric properties made the direct comparison and meta-analysis not feasible. And lastly, although the COSMIN checklist offered methodological rigor in assessing psychometric qualities of tools, it was modified by author due to inapplicability of certain criteria which were based on clinically assessed patient-related outcome measures and certain subjective judgments were made during the rating process, particularly in cases when psychometric reporting in some studies was incomplete or unclear. Such judgments, while informed by established criteria, may introduce a reviewer bias.

3.11 Conclusion

By following the PRISMA guidelines and applying the COSMIN framework for psychometric assessment, this systematic review aims to provide a narrative synthesis of FL and NL conceptualization and measurement tools among university populations. The resulting synthesis of conceptual definitions and methodological practices will offer valuable direction for researchers seeking to implement valid, reliable, and context-specific tools.

CHAPTER IV: RESULTS

The systematic literature search was conducted across multiple academic databases such as PubMed, Web of Science, CINAHL, MEDLINE, ProQuest along with additional sources such as search engines incorporating Google Scholar and websites. A total of 673 articles were retrieved at identification phase. After the de-duplication, 633 unique articles were screened based on their title and abstract. At this stage, 579 articles were excluded due to irrelevance to inclusion criteria of the study leaving 54 articles for full-text review. Following full-text screening an additional 38 articles were excluded resulting in final 16 studies included in the systematic review. The whole process of screening was reported as a flowchart according to PRISMA (Fig. 5). The reasons for the exclusion were: out of the scope of the systematic review (n=579), did not provide information on development and validation of tools (n=25), used the existing original scale without validation and adaptation (n=8), not target population (n=2), not in English (n=2), full text of the article is absent (n=1).

A total of 16 studies with 16 measurement tools assessing FL and/or NL among university students were included in this review. The tools varied in conceptual frameworks they used to develop scale, number of constructions and items, methodological approaches, and cultural adaptations. The studies were published within the 2018-2024 timeframe. 8 studies primarily targeted NL (Liao et al., 2018; Y. Liu et al., 2024; Makiabadi et al., 2019; McNamara et al., 2022; Mo et al., 2022; Mostafazadeh et al., 2024; Vrinten et al., 2023; Yan et al., 2023), 6 studies focused on FL (Durmus et al., 2019; Guiné et al., 2023; Y. Lee et al., 2022; Luque et al., 2022; Na & Cho, 2020; Rhea et al., 2020), 1 measured critical component of NL (McNamara et al., 2020), and 1 study incorporated both FL and NL measurements (Demir & Özer, 2022). The measurement tools were developed or validated across North America: USA (McNamara et al., 2020, 2022; Rhea et al., 2020), Asia: South Korea (Y. Lee et al., 2022; Na & Cho, 2020), China (Y. Liu et al., 2024; Mo et al., 2022; Yan et al., 2023),

Taiwan (Liao et al., 2018), Middle East: Iran (Makiabadi et al., 2019; Mostafazadeh et al., 2024), Turkey (Demir & Özer, 2022; Durmus et al., 2019), Europe: Portugal (Guiné et al., 2023), Spain (Luque et al., 2022), Belgium (Vrinten et al., 2023). While some measurement instruments were newly developed (Demir & Özer, 2022; Guiné et al., 2023; Liao et al., 2018; McNamara et al., 2022; Mo et al., 2022; Na MS et al., 2020; Rhea et al., 2020; Vrinten et al., 2023; Yan et al., 2023) to suit specific purposes and cultural contexts, some of them were adapted from the existing tools (Durmus et al., 2019; Y. Lee et al., 2022; Y. Liu et al., 2024; Luque et al., 2022; Makiabadi et al., 2019; McNamara et al., 2020; Mostafazadeh et al., 2024). The study by Yan et al. (2023) was added to the systematic review as a newly developed tool despite the fact that the original tool exists because the original version is in Chinese. The narrative synthesis part of the systematic review organizes the findings of the study into the following thematic domains: (1) conceptual frameworks, (2) item structure and constructs, (3) development and validation methods, (4) sample characteristics, (5) cross-cultural adaptation. These thematic areas will provide a comparative lens for identifying possible strengths and weaknesses or research gaps in development of the tools in relation to their scope and reliability.

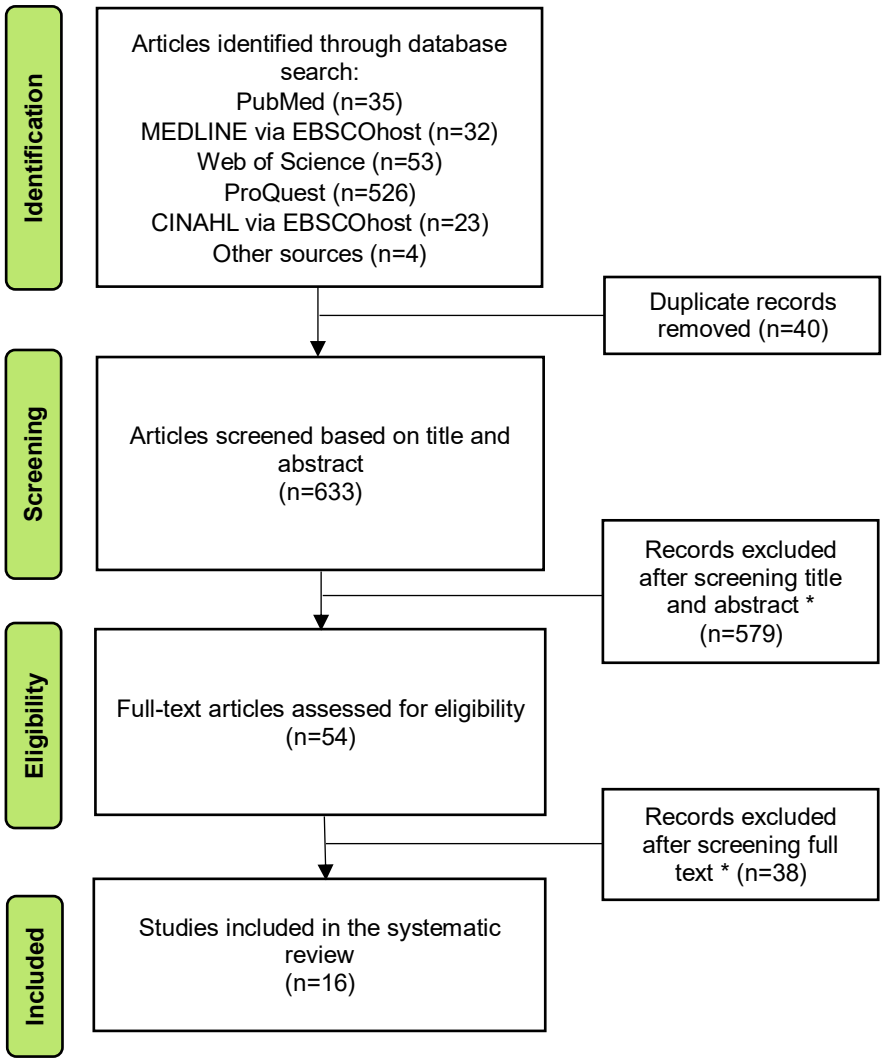

Figure 5. PRISMA flowchart

∗The reasons for the exclusion were: out of the scope of the systematic review (n=579), did not provide information on development and validation of tools (n=25), used the existing original scale without validation and adaptation (n=8), not target population (n=2), not in English (n=2), full text of the article is absent (n=1).

4.1 Conceptual frameworks

The reviewed scales mainly adopted Nutbeam's (2000) tripartite health literacy model and Vidgen & Gallegos' (2014) framework with full or partial adoption as well as with some modifications reflecting the differing interpretations of FL and NL concepts in the field of nutrition. Nutbeam's (2000) tripartite health literacy model was fully or partially adopted by eight tools: NLI-28 (Makiabadi et al., 2019), CNLT-R (McNamara et al., 2020), NL-SF12 (Mo et al., 2022), YA-NLT (McNamara et al., 2022), NLAQ (Yan et al., 2023), S-NutLit

(Vrinten et al., 2023), Iranian NL-SF12 (Mostafazadeh et al., 2024), Chinese S-NutLit (Y. Liu et al., 2024). According to this model, literacy comprises of functional, interactive, and critical literacy domains and its usage is prevalent among NL tools. This indicates that Nutbeam's model is not only well-known but also widely used as a reference model for measuring FL and NL.

In contrast, FL tools were mainly based on Vidgen & Gallegos' (2014) framework which defines FL as a set of knowledge, skills, and behaviors to plan and manage, select, prepare and eat foods to meet nutritional needs. This framework was used in scales such as KFLS (Na & Cho, 2020), EFLBQ (Rhea et al., 2020), Spanish and Korean SPFL (Luque et al., 2022; Y. Lee et al., 2022), FNLI (Demir & Özer, 2022). FNLI because of the integrated measurement of FL and NL used both Vidgen & Gallegos' (2014) framework for subject dimensions and Truman et al., (2019) literacy mapping for domain dimensions such as knowledge, attitudes, and behaviors providing a multi-dimensional approach.

A few scales adopted hybrid or modified models. The Turkish SFLQ (Durmus et al., 2019) is guided by the same conceptual frameworks as the original SFLQ (Gréa Krause et al., 2018) by Nutbeam's health literacy model along with inclusive FL concept by Krause et al. (2016), while the Portuguese FLS developed by Guiné et al. (2023) proposed a tripartite model of FL about nutritional composition, labeling and food choices, and healthy eating practices—although it did not reference a formal theoretical model. One tool, NL-TCS (Liao et al., 2018) lacked an explicitly stated conceptual framework but appeared to be based on the health literacy definition by the US National Institute of Medicine where 'health' is replaced by 'nutrition' (Nielsen-Bohlman et al., 2004).

Overall, the presence of different conceptual frameworks and their modifications emphasizes the dynamic and diverse nature of FL and NL as constructions. While Nutbeam's (2000) and Vidgen & Gallegos' (2014) models continue to dominate the field, emerging tools

are introducing integrated or adapted frameworks customized to specific cultural or behavioral contexts, emphasizing the need of contextually-driven approaches in tool development.

4.2 Item structure and used constructs

The scales demonstrated substantial heterogeneity in terms of item numbers and the breadth of constructs assessment. The number of items ranged from as few as **7** (CNLT-R) to as many as 53 (YA-NLT) items, reflecting differences in the depth and scope of literacy assessment. Short-form tools such as NL-SF12 (Mo et al., 2022), S-NutLit (Vrinten et al., 2023) were designed in a way to simplify and make the administration easier while maintaining an adequate level of psychometric characteristics by selecting items with a high factor loadings and content validity. However, the tradeoff in scale simplification may compromise reliability and validity by leading to loss of information. Tools generally measured FL and NL as constructs comprising multiple domains, often grouping items into this domains. In the case of NL tools, these domains were mainly derived from Nutbeam's (2000) tripartite health literacy model—functional, interactive, and critical NL—as measured explicitly in YA-NLT and NLI-28. For example, YA-NLT employed subscales with three domains encompassing 53 items: 6 for functional, 11 for interactive, and 36 for critical literacy, the latter further subdivided into four latent factors. Similarly, NLI-28 distributed its 28 items across the same three NL domains, ensuring conceptual alignment. Other NL tools also were informed by Nutbeam's (2000) model but dimension names and item numbers are varied. Item development of NL-TCS was informed by previous nutrition-specific literacy literature and refined through a Delphi survey of 28 experts to yield 8 self-rated items and 32 scenario-based test items utilizing 5 domains of NL: obtain, understand, analyze, appraise, and apply. Six dimensions of NL-43 (Zhang et al., 2022) were used in shortened version, NL-SF12, such as knowledge, understanding, obtaining skills, applying skills, interactive skills,

and critical skills with 2 items for each dimension. S-NutLit covered two subscales - “information skills” and “expert skills”-though scale clearly draws from both theoretical constructs and adaptations of earlier instruments. NLAQ covered three domains: Obtaining Information: 4 items; Understanding Information: 6 items; Evaluation and Application of Information: 3 items. The CNLT-R concentrated solely on critical nutrition literacy and kept only the claims-based subscale from the original Norwegian tool, yielding a concise 7-item assessment.

In the domain of FL, tools often utilized skill-based and behaviorally oriented structures. For instance, EFLBQ (Rhea et al., 2020) organized 19 items into five factors where three of them behavior-oriented such as health and nutrition, taste, and convenience, while KFLS by Na & Cho (2020) tool categorized its 50 items into knowledge and skills domains, with sub-dimensions such as food systems knowledge, self-efficacy, and food choice behavior. The FNLI (Demir & Özer, 2022), which measures both FL and NL, employed a more granular approach by integrating subject dimensions (e.g., planning, selection, preparation, eating) with domain dimensions (knowledge, attitude, behavior), yielding a total of 36 items across three domains. Guiné et al. (2023) designed the FLS, which comprised of 26 items organized into three domains: nutritional literacy, product labeling and choices, and healthy eating practices. The final model was validated as a second-order factor structure, indicating a global food literacy construct supported by interconnected subdomains. Among the updated tools, the Spanish and Korean versions of SPFL showed substantial structural changes. Following CFA, the Spanish SPFL was reduced to 26 items divided into five redefined factors, whilst the Korean SPFL retained 27 items but highlighted sustainability-oriented constructs by empirical reclassification.

The scores in all instruments were calculated using the Likert-type scales ranging from 4 to 7-point options. The scoring systems were typically summative, with some

instruments offering domain-specific subscores which can be used separately (YA-NLT, McNamara et al., 2022).

4.3 Development and validation methods

The reviewed tools varied from each other by development methods and used methodological techniques, but mainly, all of them shared a systematic, multi-stage process involving item generation, content validation, pilot testing, and psychometric evaluation. They were either newly developed or adapted from pre-existing instruments, depending on the research context. Newly developed tools started from item generation based on literature review and expert input, face/content validity assessments through Delphi panels or CVI scoring, and statistical evaluation using EFA and CFA. In turn, adapted tools used Brislin's back-translation method or WHO translation protocols, followed by cultural adaptation, item refinement, and local psychometric validation. While methods of development varied in length, complexity, and coverage of constructs, all of them demonstrated a commitment to ensure both conceptual clarity and empirical reliability.

The YA-NLT (McNamara et al., 2022) employed a six-stage development procedure which started from focus groups and literature review to generate items according to Nutbeam's (2000) model. The 133 initial items of YA-NLT were refined through expert review and psychometric validation, yielding a 42-item scale validated through EFA, CFA, and IRT. The KFLS (Na & Cho, 2020) scale item generation after literature reviews and Delphi panels produced 121 items, which later refined to an 8-factor structure. The FNLI (Turkey) (Demir & Özer, 2022) used a three-phase, nine-step procedure grounded in Vidgen & Gallegos (2014) and Truman's literacy map, beginning with item development and expert validation (CVR/CVI), followed by EFA, CFA, and known-group testing. The FLS from Portugal (Guiné et al., 2023) generated a 50 items at item generation stage across three

cognitive domains such as information access, comprehension, and application. The scale was reduced to 26 items and confirmed validity through second-order factor analysis after expert validation and pilot testing.

Other new tools such as the S-NutLit from Belgium (Vrinten et al., 2023) applied a six-phase development cycle, including expert reviews (CVI), cognitive interviews, and pre-testing. It finalized an 11-item tool with two subscales ('information skills' and 'expert skills') validated through EFA, CFA, and test-retest reliability. The NL-TCS (Liao et al., 2018) involved Delphi panels and a dual-format test (8 self-rated and 32 scenario-based items) grounded in five NL dimensions. The NLAQ (Yan et al., 2023) was built on local nutrition education theory and validated using EFA, CFA, and LPA. It comprises 13 items across three domains: obtaining, understanding, applying information. The NL-SF12 (Mo et al., 2022) was a shortened version of NL-43 (Zhang et al., 2022), selected using factor loadings and regression to retain two items per six domains. Its Iranian adaptation by Mostafazadeh et al. (2024) replicated the procedure with local validation.

Adapted tools followed translation and psychometric protocols. The CNLT-R (McNamara et al., 2020), revised version of a Norwegian tool, retained only the "claims" subscale and excluded the "engagement" scale for contextual relevance in U.S. university settings. Final validation yielded a 7-item scale. The Spanish SPFL (Luque et al., 2022) underwent translation, back-translation, pilot testing, and CFA. Items related to social eating were removed, resulting in a 26-item, five-factor model with high reliability. The Korean SPFL (Y. Lee et al., 2022) retained 27 items but restructured factor groupings around sustainability and behavioral outcomes, with validation via CFA and SEM. The Turkish SFLQ (Durmus et al., 2019) was adapted from the Swiss SFLQ (Gréa Krause et al., 2018) using translation/back-translation, followed by expert review (CVI = 0.67) and validation through EFA and test-retest reliability. A one-factor structure was retained, explaining 32% of

variance. Finally, the CS-NutLit (Y. Liu et al., 2024) followed Brislin's translation protocol and expert review. Local adaptations included replacing culturally specific examples and validating the tool through EFA, CFA, and test-retest procedures (ICC = 0.818).

Across all instruments, item generation was based on theoretical knowledge in the field and the context of study. Content validation was universally performed using expert panels, and pretesting was common through cognitive interviews or pilot studies. Psychometric validation relied on EFA and CFA, with several tools incorporating IRT (e.g., YA-NLT, CNLT-R), SEM (e.g., EFLBQ, Korean SPFL), and test-retest reliability (e.g., FNLI, S-NutLit, Turkish SFLQ, KFLS, EFLBQ, Iranian NL-SF12, CS-NutLit). Overall, the measurement tools reflect an evolving field characterized by methodological rigor and increasing attention to cultural and behavioral specificity in FL and NL assessment.

Table 1: A summary table of FL and NL tools for university students

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
Liao et al., (2018)	Taiwan	NL	NL-TCS	To develop a nutrition literacy measure for Taiwanese college students and evaluate its reliability and validity	Not specified, assumed to be based on the Health Literacy definition by the US National Institute of Medicine where 'Health' is replaced by 'Nutrition'.	8 self-rated items and 32 scenario-based test items utilizing 5 domains of NL: *obtain, understand, analyze, appraise,* and *apply*	4 stages: 1. Development of potential indicators; 2. Delphi survey and selection of indicators; 3. Test development; 4. Pretesting	Likert scale: from 1 (very difficult) to 4 (very easy), sum of each item score is indicator of self-rated NL score, higher scores mean better NL	Sample included (n=1269) students: women slightly outnumbered men (men-to-women ratio 0.87), with sophomores (38.1%) and juniors (26.1%) being the most common. Most participants had parents with at least a high school education (90.3%), about half lived in dorms (52.5%), and only 15.6% lived with family.
Durmus et al., (2019)	Turkey	FL	Turkish SFLQ	To translate and adapt the SFLQ into Turkish and evaluate its validity and reliability for	Guided by inclusive FL concept by Krause et al. (2016) and Nutbeam's (2000)	12 items covering functional, interactive, critical FL (Krause, Sommerhalder,	Two stage development: 1. Translation & Content validation: The translation-back translation method was used. A	Four- or five-point Likert scale with scores ranging from 7 to 52.	Convenience sampling: 308 students with age range 18-30 years (Mean: 19.94 ± 2.42 years). 28.2%

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				use among university students in Turkey.	tripartite HL conceptual frameworks	& Beer-Borst, 2016a; Krause, Sommerhalder, Beer-Borst, et al., 2016b)	panel of 10 experts evaluated content validity. 2. Validity and reliability testing: Face validity (n=10), test-retest reliability (n=40).		Male, 71.8% Female.
Makiabadi et al., (2019)	Iran	NL	NLI-28	To assess the reliability and validity of the Persian version of the Nutrition Literacy Inventory (NLI-28) by Bari (2012) among university students	Based on Nutbeam's (2000) tripartite health model	28 items designed to measure the three subscales of nutrition literacy—9 items for FNL, 9 for INL, and 10 for CNL.	Initially, the NLI-28 was translated into Persian following the WHO's guidelines for translation and back-translation. Face and content validity were established through expert review, where content validity ratio (CVR) and content validity index (CVI) values were calculated, showing high levels of agreement among experts. The instrument was then pre-tested on a small group of students before	A 5-point Likert scale for each item. For positive items, scores range from 5 (strongly agree) to 1 (strongly disagree), with negative items reverse scored so that higher overall scores reflect greater nutrition literacy.	The study sample consisted of 203 university students (101 women and 102 men) from Shiraz University of Medical Sciences, with a mean age of 22.10 ± 1.89 years. Participants were selected through a cluster random sampling method from four dormitories, ensuring that the sample met the required item-to-participant ratio for robust factor analysis.

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
							being administered in the main study. The psychometric evaluation involved both EFA and CFA, as well as reliability testing using Cronbach's alpha and split-half methods	The total score ranges from 28 to 140.	Demographic data such as weight, height, and academic discipline were also collected.
McNamara et al., (2020)	USA	CNL	CNLT-R	To refine and psychometrically validate the Revised CNLT in a young adult sample	Nutbeam's (2000) tripartite health model with emphasis on CNL.	7 items distributed across two distinct constructs: Factor 1: Critical appraisal of media sources (4 items); Factor 2: Critical appraisal of evidence-based nutrition sources (3 items)	Refinement of CNLT: Revised for U.S. young adults, with wording adjustments for relevance. Psychometric Validation: EFA and CFA assessed structure; IRT evaluated item difficulty and discrimination. Reliability Testing: Measured internal consistency using Cronbach's alpha (α) and coefficient omega (ω).	Likert scale: from 1 (strongly disagree) to 5 (strongly agree)	The sample included 1,718 young adults recruited from three geographically diverse universities in the United States. Demographic details: Age: Average around 20 years; Gender: Approximately 70% female; Ethnicity: Predominantly white (81%); Living Situation: A majority lived

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
									off campus (67%).
Na & Cho, (2020)	South Korea	FL	KFLS	To develop and validate a food literacy assessment tool specifically for young Korean adults	Multiple theoretical models: Vidgen & Gallegos (2014), scoping review-based model categorizing FL into 11 components by Azevedo Perry et al. (2017)	50 items divided into two domains: Knowledge Domain (25 items): *Food and nutrition knowledge* (11 items); *Food safety* (4 items); *Food systems* (4 items); *Sociocultural context* (6 items). Skills/Ability Domain (25 items): *Food skills*; *Self-efficacy*; *Food choice*; *Food resource management*.	1.Development and Content Validation: generation of 121 items covering 11 FL components. A Delphi survey with 15 experts; 2.Pilot Study: The revised tool was pilot tested with 108 young Korean adults via an online survey using a 5-point Likert scale. Internal consistency was evaluated using Cronbach's α, ensuring each item fit its assigned construct. 3.Confirmatory Study.	A 5-point Likert scale. Sum scores for each domain are calculated, with higher total scores indicating higher food literacy.	Sample included 543 young adults. Pilot Study Sample: 108 young Korean adults (aged 20-29) Confirmatory Study Sample: 435 young Korean adults (aged 20–29) with a mean age of 24.6 ± 2.6 years. Gender distribution: Approximately 44.4% male and 55.6% female. Educational attainment ranged from high school to bachelor's degree (majority at the bachelor's level).
Rhea et al., (2020)	USA	FL	EFLBQ	To develop and validate a questionnaire	Framework by Vidgen & Gallegos (2014)	The CFA supported a 5-factor model comprising:	Item Generation and Initial Testing (Study 1) Refinement and	A 4-point Likert-type scale with response	Sample: 1,822 students. Mean age: ~20.8 years. Gender: ~64%

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				that measures university students' perceived food literacy behaviors.		Factor 1: Health and nutrition (7 items); Factor 2: Taste (3 items); Factor 3: Food preparation (3 items); Factor 4: Planning and/or decision-making (3 items); Factor 5: Convenience (3 items) This final model includes a total of 19 items.	Second Administration (Study 2); Final Validation (Study 3): (EFA) with 265 students and then (CFA) with 929 students from multiple courses. Test–retest reliability was assessed on a subsample of 67 students, and (SEM) was used to develop an explanatory model for the relationships among the identified factors.	options: 1 = never - 4 = always, with an additional "does not apply" option. Higher scores indicate a greater frequency of the healthy food literacy behaviors.	female. Ethnicity: ~76% White/Caucasian
Mo et al., (2022)	China	NL	NL-SF12	To develop and validate a short-form version of the NL-43 for Chinese college students which name is NL-SF12.	Based on a multi-level health literacy framework by Nutbeam (2000), integrating functional, interactive, and critical	The newly developed NL-SF12 includes 12 items in total—specifically, 2 items per each of the six dimensions: *knowledge*, *understanding*,	1.Data Collection and Partitioning; 2.Exploratory Factor Analysis (EFA): EFA with oblique Promax rotation was used on the first dataset to extract factors based on eigenvalues ≥1 and	A 5-point Likert scale (1 = strongly disagree - 5 = strongly agree) for each item. Total scores are computed	1359 Chinese college students from Anhui Province Demographics: Mean age: 22.22 ± 2.94 years Gender: 40.7% male and 59.3% female

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
					nutrition literacy to assess comprehensive nutrition understanding inspired by Velardo (2015).	*obtaining skills*, *applying skills*, *interactive skills*, and *critical skills*.	factor loadings >0.40. 3.Item Selection via Regression Analysis: Items were further refined by creating subsets based on the highest factor loadings. Linear regression models were used—with the full NL-43 score as the dependent variable—to select the best-performing items. 4.Validation and reliability testing.	by summing responses across the 12 items, with higher scores indicating higher nutrition literacy.	Academic Level: Freshmen (26.2%), Sophomores (37.1%), Juniors (16.9%), and Seniors (19.8%) Type of College: Approximately 58% were enrolled in medical colleges and 42% in non-medical institutions Response Rate: 96.5%
Luque et al., (2022)	Spain	FL	Spanish SPFL	To adapt and validate a Spanish version of the SPFL Scale by Poelman et al. (2018) for use with university students.	The original SPFL is based on framework by Vidgen & Gallegos (2014)	16 items which grouped into five factors: *Cooking skills*, *Emotional management*, *Healthy consumption as a priority*, *Nutritional literacy and*	1.Translation and Cultural Adaptation; 2.Preliminary Testing with university students; 3.Validation and reliability testing.	A 5-point Likert-type scale ranging from 1 ("never") to 5 ("always"). Reverse scoring is applied to designated	The validation study was conducted with 362 Spanish university students from diverse degrees and institutions. Participants' age ranged from 18 to 36 years, with a mean age of

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
						planning, *Availability of ultra-processed foods*.		items so that a higher total score consistently indicates a higher level of food literacy.	22.36 ± 3.76 years. The sample was predominantly female (approximately 86.7%).
Lee, Y. et al., (2022)	South Korea	FL	Korean SPFL	To examine whether university students' perceived food literacy influences their ecological eating behavior, which is defined as eating in a way that minimizes environmental harm and supports sustainability	The original SPFL is based on framework by Vidgen & Gallegos (2014)	27 items assessing various dimensions such as *Food Preparation Skills, Resilience and Resistance, Healthy Snack Styles, Social and Conscious Eating, Healthy Food Stockpiling, Reading Labels and Budgeting*.	1.Adaptation and Translation: The instrument was adapted and then translated into Korean using Brislin's back-translation procedure to ensure cultural and linguistic appropriateness. 2. Development and Validation of the Measurement Tool; 3. Preliminary Testing; 4. Statistical Validation of the SPFL Scale.	A 7-point Likert scale: participants rated each item to indicate the degree of their agreement, with higher scores generally representing higher levels of food literacy or a greater tendency toward ecological	395 university students in Seoul, South Korea. Demographic Profile: Male: 41.8% Female: 58.2% Grade Level: Freshmen: 17.0% Sophomores: 22.8% Juniors: 25.1% Seniors: 35.2% Education Level: Community college (2-year degree): 12.7% University (4-year degree): 87.3% Major: Food-related majors: 28.1%

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
								eating behavior.	Other majors: 71.9%
McNamara et al., (2022)	USA	NL	YA-NLT	To develop and validate an instrument—the Young Adult Nutrition Literacy Tool (YA-NLT)—to assess college students' perceived nutrition literacy across the functional, interactive, and critical domains.	Nutbeam's, (2000) tripartite health model	FNL: 6 items, 2 factors INL: 1 factor, 11 items. CNL: 36 items, 4 factors.	1. Focus group: item generation; 2. Expert review. 3. Exploratory factor structure analysis. 4. Item refinement and modification. 5. Factor structure validation; 6. Criterion validation.	All items are rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).	Sample included 672 students. Demographics (Validation Sample): Gender: Approximately 62.5% female. Race/Ethnicity: Predominantly White (82.5%). Age: Mean age was 19.9 years (± 1.8 years). Dietary Intake: Average fruit and vegetable intake was 2.2 servings per day. Recruitment Methods: Convenience sampling.
Demir & Özer (2022)	Turkey	FL and NL	FNLI	To develop a valid and reliable measurement tool that can evaluate FNL	Framework by Vidgen and Gallegos (2014) in the subject dimensions and the	Subject dimensions were determined as *planning and management*, *selection*,	Phase 1: Item Development Step 1: Items generated from an extensive literature review (Turkey Nutrition Guide 2015,	Scoring by domain: Knowledge domain – score range 0-13 points; Attitude	The instrument was administered to students aged 18–21 at X University during the spring semester of the

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				holistically in young people	literacy mapping of (Truman et al., 2019) in the domain dimensions were used	*preparation*, and *eating*, and domain dimensions were determined as *knowledge*, *attitudes*, and *behaviors* corresponding to *declarative*, *procedural*, and *subjective* knowledge. The final version consists of 36 items: Knowledge Domain: 13 items Attitude Domain: 13 items. Behavior Domain: 10 items.	Vidgen & Gallegos, Truman et al.). Step 2: Content validity assessed by 13 experts using Lawshe's technique. Phase 2: Scale Development Step 3: Pre-testing conducted with 20 students. Step 4: Quota sampling of 538 university students. Step 5: Items reduced using correlations, difficulty, and discrimination indices. Step 6: EFA confirmed factor structure. Phase 3: Scale Evaluation Step 7: CFA confirmed model fit. Step 9: Validity confirmed via "known groups" analysis.	domain – score range 13-65 points; Behavior domain – score range 10-50 points.	2019–2020 academic year. Total Valid Responses: 538 young people. Gender: 57.6% female (n = 310) 42.4% male (n = 228) Age: Mean age: 19.2 ± 0.9 years Median: 19.0 years (range: 18–21 years). Data were collected using paper and pencil interviews (PAPI) during May 2020. Quota sampling was used to ensure a sex-specific proportional representation from the total population of 4359 students.
Guine et al., (2023)	Portugal	FL	FLS	The primary goal of the study was to	Particular conceptual framework	26 items distributed across three	1.Item Development and Content Validity	All items are rated on a 5-point	Sample included 924 Portuguese higher education

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				test and validate a Food Literacy Scale (FLS) for Portuguese university students.	not mentioned but the model where food literacy is viewed as a multidimensional construct that encompasses literacy about the nutritional composition of foods; literacy about food labelling and food choices; literacy about healthy eating practices	factors: F1 – Literacy about the Nutritional Composition of Foods (10 items); F2 – Literacy about Labelling and Food Choices (7 items); F3 – Literacy about Healthy Eating Practices (9 items)	2.Scale Development: Pre-testing: A pre-test phase (with a small number of participants) was conducted to ensure that items were clear and adapted to the target population. Validity and reliability assessments: EFA, CFA, internal consistency analysis	Likert scale, where "Very easy" (4) indicates higher literacy, and lower scores (closer to 1) indicate lower literacy. Responses of "I do not know" (5) are analyzed separately.	students (university/polytechnic). The recruitment followed a convenience sampling method with quota sampling. Age: 18–57 (F), 18–70 (M); Mean age: 21.9 (F) / 24.1 (M) Main Age Group: ≤19 yrs (F), ≥22 yrs (M) Degree Level: Licence: 62.1% (F) / 12.9% (M); Master: 13.3% (F) / 4.3% (M)
Yan et al., (2023)	China	NL	NLAQ	To evaluate the reliability and validity of the Nutrition Literacy Assessment Questionnair	The framework is informed by Nutbeam's health literacy model, but the NLAQ	The NLAQ comprises a total of 13 items divided into three domains: Obtaining Information: 4	1.Item Development; 2.Questionnaire Design: two-part questionnaire with part 1 gathering demographic, academic, and	Items are scored on a five-point Likert scale, with responses ranging from 1	Sample included (n=774): 57.5% male, age range 18–23 years. Academic Background: 66.93% had >3 hours of online

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				e (NLAQ) for college students and to identify the influencing factors of their nutrition literacy.	focuses on stage-specific CNL aspects relevant to college students (Guttersrud et al., 2014).	items; Understanding Information: 6 items; Evaluation and Application of Information: 3 items.	family environment information, and part 2 focusing on nutrition literacy through the 13 items. 3.Pre-testing; 4.Psychometric Evaluation; 5.Additional Analysis: Latent profile analysis (LPA) was used to classify nutrition literacy levels, and chi-square tests with logistic regression were conducted to identify influencing factors.	(totally disagree) to 5 (totally agree). Higher scores indicate higher nutrition literacy.	learning, 51.7% sophomores, 49.6% science and engineering students. Family Environment: 28.04% from county-level cities, 37.08% annual household income 50,000–100,000 CNY. Parental Background: Majority had junior high school education or below, with employment in small businesses or general staff roles.
Vrinten et al., (2023)	Belgium	NL	S-NutLit	To develop and validate a short nutrition literacy scale (S-NutLit) tailored for young adults aged 18–25 years.	Nutbeam's (2000) tripartite HL model.	11 items: 'Information Skills' - 8 items assessing abilities such as seeking, applying, and appraising nutrition	1.Item Generation; 2.Expert Assessments; 3.Cognitive Interviews; 4.Pre-test; 5.Validation Survey; 6.Test–Retest Reliability Survey;	A 5-point Likert scale: 1 = "completely disagree" 5 = "completely agree". Higher scores	Cognitive Interviews (n = 12): Mean age 22.6 years, 50% women. Pre-Test Survey (n = 101): University students. Validation Survey (n =

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
						information. 'Expert Skills' - 3 items measuring reliance on and use of expert or scientific nutrition information.		indicate a higher level of nutrition literacy.	300): Recruited online (ages 18–25, 53% women, 28% health/nutrition field). Test-Retest Reliability (n = 92): Conducted two weeks later, 67.4% women.
Mostafazadeh et al., (2024)	Iran	NL	Iranian NL-SF12	To assess the relationship between nutrition literacy (NL) and eating behaviors among nursing students at Ardabil University of Medical Sciences, Iran	Nutbeam's (2000) health model	12 items covering 6 domains: *Knowledge, Understanding, Obtaining Skills, Applying Skills, Interactive Skills,* and *Critical Skills*	1. Adaptation from the Chinese NL-SF12; 2. Translation and back-translation (from English to Persian and back); 3. Content Validity Testing: Evaluated by 10 faculty experts; 4. Factor analysis (EFA, CFA)	5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Higher scores indicate greater nutrition literacy	Sample included 224 nursing students (98 males, 126 females). Mean age = 22.8 ± 3.16 years. Sampling Method: Proportional stratified random sampling. Exclusion: Those with neurological/psychiatric disorders or incomplete responses.
Liu et al., (2024)	China	NL	CS-NutLit	To translate, culturally adapt, and validate the	Nutbeam's (2000) tripartite HL model	11 items, divided into two factors:	3 phase: 1. Translation & Cross-cultural validation; 2. Item	Likert scale with scores ranging from 1 to 5	Convenience sampling: 508 young adults. Age Groups: 18-

Author (year)	Country of origin	Measured construct	Tool name	Purpose	Conceptual Framework	Number of Items & Constructs	Development method	Scoring details	Sample characteristics
				S-NutLit tool for the young Chinese adults.		1.Information Skills; 2. Expert Skills	analysis 3. Validation analysis.	with an additional answer option for item 7 which is not included in total score.	25 years old: 60.6%. 26-35 years old: 39.4%. Gender: Male: 42.3%. Female: 57.7%. Education Level: Middle school or below: 3.9%. High school or junior college: 21.3%. Undergraduate: 61.6%. Postgraduate: 13.2%.

NL-TCS, Nutrition Literacy for Taiwanese College students; Turkish SFLQ, Turkish Short Food Literacy Questionnaire; NLI-28, Nutrition Literacy Inventory – 28 items; CNLT-R, Critical Nutrition Literacy Tool – Revised; KFLS, Korean Food Literacy Scale; EFLBQ, Eating and Food Literacy Behaviors Questionnaire; NL-SF12, Nutrition Literacy – Short Form 12 items; Spanish SPFL, Spanish Self-perceived Food Literacy; Korean SPFL, Korean Self-perceived Food Literacy; YA-NLT, Young Adults – Nutrition Literacy Tool; FNLI, Food and Nutrition Literacy Instrument; FLS, Food Literacy Scale; NLAQ, Nutrition Literacy Assessment Questionnaire; S-NutLit, Short-Nutrition Literacy; Iranian NL-SF12, Iranian Nutrition Literacy – Short Form 12 items; CS-NutLit, Chinese Short – Nutrition Literacy.

4.4 Sample characteristics

The samples used to validate the FL and NL tools differed greatly in size, demographic composition, academic background, and recruitment strategies. Despite this difference, all studies focused on university students or young adults whose age coincides with education time at university, which is consistent with the population of interest for this study. Sample sizes in reviewed articles ranged from 203 (NLI-28, Makiabadi et al., 2019) to 1,822 (EFLBQ, Rhea et al., 2020) participants, with the majority of studies enrolling approximately 300 to 1,000 participants, which was considered as sufficient for validation and reliability testing.

It was found that gender distribution skewed more towards female participants, with their proportions ranging from ~57% to over 85% in several studies (Spanish SPFL by Luque et al., 2022 and FLS by Guine et al., 2023). This disparity could be attributed to women's overrepresentation in health and nutrition-related fields, as well as female students' stronger interest in food-related habits and behaviors. However, it adds the risk of gender bias in item responses and restricts the generalizability of findings to mixed-gender populations unless gender invariance is explicitly evaluated which in the following studies was not done.

Most reviewed studies involved participants aged 18 to 25, with some extending to 30 or beyond (Turkish SFLQ by Durmus et al., 2019 and Spanish SPFL by Luque et al., 2022). The average participant age across studies ranged between 19 and 23 years, representing usual university level demographics. A few tools, such as FNLI (Demir & Özer, 2022) and S-NutLit (Vrinten et al., 2023), strictly limited participation to people aged 18 to 25, which improved internal homogeneity but reduced applicability across older students or postgraduate cohorts.

Academic background of participants were reported inconistently, as some studies provided a more details about the disciplines by grouping participants into medical vs non-

medical (NL-SF12), science related disciplines (NLAQ, Spanish SPFL), food and nutrition-related majors (Korean SPFL, S-NutLit) demonstrating that FL and NL may differ based on field of study. NLI-28 was validated among nursing students and other health-related specialties. Other tools, such as the Turkish SFLQ, EFLBQ, YA-NLT, FNLI, FLS, focused on broad representation within universities with degree levels along with BMI indeces. Socioeconomic and educational background variables including parental education, monthly income, dormitory status, urban vs. rural origin of participants were mentioned in only a few studies such as NL-TCS, KFLS, NLAQ, CNLT-R, Iranian NL-SF12, although such factors may contribute to shape FL/NL outcomes.

The majority of studies used convenience or quota sampling, with recruitment taking place via online platforms, institutional email lists, or in classrooms. Only a few studies used randomized or stratified sampling methods: NLI-28, Iranian NL-SF12, S-NutLit which are better for reducing selection bias. The lack of random sampling reduces the external validity of most instruments, yet this is compatible with widely accepted scale development practices.

To summarize, while sample sizes were being generally appropriate, participant pools were not demographically equal in terms of gender or academic background. The use of mainly convenience sampling and underreporting of socioeconomic and contextual details in some studies can limit the capacity to analyze how well each measurement instrument performs across different subpopulations. Future research could benefit from more systematically reported and demographically stratified populations, as well as examining measurement invariance across major demographic groups, to improve the equity and generalizability of FL/NL tools.

4.5 Cross-cultural adaptation

Cross-cultural adaptation criteria were applied to several scales, especially those developed in non-English contexts or adapted from pre-existing tools. Among the 16 studies included, 7 tools were explicitly adapted from existing instruments rather than being created from scratch: Turkish SFLQ (Durmus et al., 2019), Spanish and Korean SPFL (Luque et al., 2022; Y. Lee et al., 2022), Iranian NL-SF12 (Mostafazadeh et al., 2024), NLI-28 (Makiabadi et al., 2019), CNLT-R (McNamara et al., 2020) and CS-NutLit (Y. Liu et al., 2024). Some original tools, such as NL-TCS (Liao et al., 2018) and KFLS (Na & Cho et al., 2020), were also rewritten and revised according to cultural specificities during the development phases. Still, they are not validated outside of their countries, limiting their applicability across different cultures.

Spanish SPFL, CNLT-R, and Korean SPFL underwent structural changes along with linguistic adaptation, while Iranian NL-SF12, Turkish SFLQ, and NLI-28 relied mainly on back-translation with the same number of factors and items. Adapted tools usually used conventional translation and back-translation techniques such as Brislin's method, to ensure linguistic equivalence. This was frequently complemented by expert review panels to determine relevance of the content in the target culture or setting, as shown in the Turkish SFLQ, NLI-28, Korean and Spanish SPFL, Chinese S-NutLit, and Iranian NL-SF12. Several studies additionally conducted face validity tests with local students to improve culturally sensitive items and compehensibility (Korean SPFL). These procedures are good practice for ensuring cultural appropriateness. CS-NutLit deleted the Europe-specific notion of ‘Fair Trade coffee’ and reworded the education model of Belgium called ‘Flemish Food Triangle’ to its Chinese equivalent ‘Chinese balanced diet pagoda’. Pre-testing (n=50) was conducted with students to assess its clarity and comprehensibility after cultural adaptation.

Table 2: The psychometric properties of FL and NL tools for university students

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
NL-TCS Liao et al., (2018)	Delphi survey with 28 experts. (CVI): Self-rated scale: 1.00 (perfect agreement) Scenario-based test: 0.99	EFA: PCA Scree Plot: Supported a single construct. CFA: Self-rated scale: GFI = 0.993, AGFI = 0.988, RMSEA = 0.019 Scenario-based test: GFI = 0.926, AGFI = 0.913, RMSEA = 0.015. Scenario-based test difficulty analysis: Item response difficulty range: -3.22 to 0.11 (acceptable range: -4 to 2). Spearman's Rank Correlation: Self-rated scale ↔ Scenario-	Not reported	Based on: Taiwanese Dietary Guidelines and Daily Food Guides	Cronbach's alpha (Self-rated scale): 0.85 Cronbach's alpha (Scenario-based test): 0.81	Not reported	Comprehensive Psychometric Validation. Dual Assessment Approach: Self-rated scale: Measures perceived nutrition literacy. Scenario-based test: Evaluates actual ability to apply nutrition knowledge in real-world situations. Large, Representative Sample.	77.4% average correct responses in the scenario-based test suggest many items were too easy leading to ceiling effect. Convenience sampling which lowers generalizability. The scale is based on the Dietary Guidelines for Taiwanese which limits its applicability in other populations.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		based test: r = .155, p < .001 Both scales ↔ Healthy-eating behaviors: r = .417, p < .001 and r = .125, p < .001						
Turkish SFLQ Durmus et al., (2019)	Expert Review conducted with 10 experts. (CVI): 0.67	EFA: (KMO = 0.811, Bartlett's Test of Sphericity: χ^2 = 841.958, df = 66, p < 0.001, 1-factor model, 32.01% variance, Factor Loadings = 0.43 - 0.64). Correlation with Turkey Health Literacy Scale-32 (TSOY-32): r = 0.531, p < 0.001. Newest Vital Sign (NVS): r = 0.294, p < 0.001. Higher	Not reported	Translation, Back-Translation, Expert Panel, Pilot Testing	Cronbach's α = 0.803. Item-Total Correlation: r = 0.32 - 0.52. Cronbach's α if item deleted = 0.77 - 0.79.	Three-week interval test-retest (n = 48). Spearman Rank Correlation: r = 0.808, p < 0.001. Mean Score Change: First Test: 40.89 ± 6.86. Retest: 42.61 ± 7.56.	Sound psychometric characteristics, and ease of use for assessing FL of university students. Validated translation and adaptation with the participation of experts.	Lower variance explained in comparison to original scale: 32.01% vs 76.4%. Potential social desirability bias due to self-reported design of scale.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		health literacy correlated with higher SFLQ scores, supporting hypothesis-testing validity						
NLI-28 Makiabadi et al., (2019)	Expert Review: Conducted by 6 health and nutrition experts. CVR: 0.89. CVI: Simplicity: 0.90 Clarity: 0.96 Relevance: 0.98	EFA (KMO = 0.704, Bartlett's Test of Sphericity: χ^2 = 1990.0, df = 378, 4-factor model, 78% variance), CFA (RMSEA = 0.002, CFI = 0.963, GFI = 0.942, AGFI = 0.910). Convergent Validity (r = 0.74), Divergent Validity (r = -0.11)	Criterion Validity: r = 0.73 ($p < 0.01$)	Translated and back-translated following WHO protocol. Adapted for Persian-speaking students.	Cronbach's α = 0.87	Not reported	Comprehensive multidimensional structure with 4-factors and robust psychometric characteristics. Validated translation process. Novelty of the scale in the Iran and other Persian-speaking populations. The analysis of 'well-known groups' showed that the scale can differentiate well based on sex as the original scale.	Cluster random sampling within one university which may not fully represent students from various educational backgrounds and social status. Self-administered design can lead to some degrees of bias. Further study is needed with various demographic groups and larger sample size.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
CNLT-R McNamara et al., (2020)	Expert-reviewed, adapted for U.S. students from the Norwegian version	EFA (2-factor, 44% variance), CFA (χ^2/df = 12.43, CFI = 0.90, RMSEA = 0.11, SRMR = 0.07). IRT: Items differentiate between low/high CNL levels.	Not reported	Not explicitly reported, but the adaptation was mentioned.	Cronbach's Alpha (α) =0.69 and Omega (ω) = 0.79	Not reported	Sound psychometric characteristics. Large and diverse population including students from three geographically diverse U.S. universities. Scale addresses social media and misinformation as a concern for young adults, making it relevant for the digital age.	No functional/interactive literacy assessment. The sample was predominantly white and female, limiting generalizability to diverse populations. Lack of content validity testing: no cognitive interviews and ambiguity associated with media-related terms. No assessment of whether CNL correlates with dietary behaviors. Questionable internal consistency of factor 2.
KFLS Na & Cho (2020)	Delphi expert review (15 experts), CVR > 0.51	EFA (KMO = 0.924 - 0.952, 8-factor model, Knowledge Domain: 33.46%, Skills/Ability Domain: 44.91%), CFA	Not reported	The adaptation process removed or revised certain items that were not relevant to	Cronbach's α = 0.737 - 0.910	ICC = 0.90	Scale incorporates Korean dietary habits, food culture making it highly applicable for Korean young adults. Robust 8-factor	Participants recruited via online panel in Seoul and Gyeonggi province—not nationally representative. 50 items tool, which may complicate the administration

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		(CFI = 0.976, TLI = 0.975, RMSEA = 0.064)		Korean food practices			multidimensional structure with strong psychometric validation. Delphi-based content validation. Scale addresses multiple aspects of FL.	feasibility in time-constrained settings. The tool was not benchmarked against existing instruments or correlated with dietary intake measures or health outcomes, limiting external validation. The tool is highly contextualized to Korean dietary and cultural practices, lowering the applicability outside Korea.
EFLBQ Rhea et al., (2020)	Expert review, face validity: cognitive interviews with 2 students living in dormitories, 2 students living in apartments	EFA (KMO = 0.817, 5-factor model, 57.4% variance), CFA (χ^2/df = 4.209, CFI = 0.98, TLI = 0.98, RMSEA = 0.06, SRMR = 0.05). SEM confirms links between food literacy & eating behavior	Not reported	Not reported	Cronbach's α = 0.63-0.89	Pearson correlation coefficient = 0.63 (95% CI, 0.46−0.76; P < 0.001) - 0.92 (95% CI, 0.87 −0.95; P < 0.001)	Comprehensive multi-factor structure with strong psychometric characteristics. SEM demonstrated positive influence of health and nutrition knowledge on	Limited generalizability: convenience sampling of participants from a large southeastern U.S. university, with most participants being White and female. Potential educational bias as the findings may not apply to individuals

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
							food preparation and planning behaviors. Large and diverse sample. Scale includes behavioral influences, such as taste and convenience, which are critical determinants of food choices.	with a lower educational level. The Planning & Decision-Making and Convenience factors had lower reliability compared to other EFLBQ domains. Lack of quantitative content validity testing.
NL-SF12 Mo et al., (2022)	Expert-reviewed, CVI > 0.90	EFA (KMO = 0.969, 6-factor model, 71.4% variance), CFA (χ^2/df = 4.209, CFI = 0.972, TLI = 0.953, RMSEA = 0.069, GFI = 0.960). Convergent validity: AVE values ≥ 0.50 for most factors (except Applying Skills = 0.447,	Correlation with NL-43: r = 0.969 with NL-43	Not reported, developed specifically for Chinese college students.	Cronbach's α = 0.890. Subscale Reliability (Spearman-Brown Coefficient) = 0.589-0.890	Not reported	The short-form scale accurately reflects the original scale. Efficient and practical short-form measurement with strong psychometric validation and high internal reliability. Culturally relevant to Chinese population.	Limited external validity: the scale developed and validated only on Chinese college students in Anhui. Tradeoff in scale simplification: reducing the number of items can increase the ease of administration but may compromise reliability and validity by leading to loss of information. Obtain and Apply

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		slightly below threshold). CR values ≥ 0.60, indicating strong construct reliability.						subscales showed weak convergent and discriminant validity.
Spanish SPFL Luque et al., (2022)	Expert-reviewed, translation & back-translation	EFA (5-factor model, better fit than original 7-factor model), CFA (CFI = 0.95, TLI = 0.94, RMSEA = 0.045, GFI = 0.91). Convergent (r = 0.59 with Mediterranean Diet, Self-Perceived Health (r = 0.31, $p < 0.01$). Divergent (r = -0.26 with Impulsivity)	Not reported	Adapted and translated for Spanish university students	Cronbach's α = 0.894	Not reported	Comprehensive validation against health and diet measures. Addresses psychological factors such as impulsivity along with traditional FL components. Strong psychometric validation.	Limited generalizability: despite the sufficient sample for validation, it may not fully represent the Spanish university population and low male representation. The study covers the main 4 domains of FL according to Vidgen & Gallegos (2014) instead of all 11 components which can increase the accuracy.
Korean SPFL Lee, Y. et al., (2022)	Expert-reviewed, translation & back-translation.	CFA (6-factor model, χ^2 = 1552.589 (df = 474), $p < 0.001$, NFI = 0.860,	Not reported	Validated and adapted for South Korean university students but	Cronbach's α = 0.840–0.929. CCR = 0.795-0.888	Not reported	Strong methodology, novel focus on sustainability, and	Self-reported administration, which can lead to social desirability bias. The study was

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
	Face validity: Preliminary university students to refine unclear items	CFI = 0.90, TLI = 0.89, RMSEA = 0.08). All factor loadings (λ) were ≥ 0.60, supporting convergent validity. (AVE) ≥ 0.50, confirming discriminant validity. Convergent validity: correlation with Ecological Eating Behavior scale. SEM demonstrated significant relationships for Healthy Snack Styles (β = 0.477, p < 0.001), Resilience & Resistance (β = 0.232, p < 0.001), Healthy Food		no explicit cross-cultural validity testing			comprehensive food literacy model with 6-factors. Applicable for FL and sustainability studies.	conducted only on university students in Seoul, making it difficult to apply findings to broader populations. There is a need to consider individual variables such as sensitivity to FL and cultural differences to better understand variations in food behavior.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		Stockpiling (β = 0.198, p < 0.01), and Labels & Budgeting (β = 0.134, p < 0.05).						
YA-NLT McNamara (2022)	Focus Groups (n = 24 college students) for item generation. Expert Panel Review (n = 10) in nutrition and dietetics assessed item importance. Item refinement and classification by the research team.	EFA (multi-factor) and CFA: FNL: 2 factors - excellent model fit (CFI = 0.99, RMSEA = 0.04, SRMR = 0.03) Strongest CFA Loading: 0.97, IRT Slope: 6.05. INL: 1 factor - Moderate model fit (CFI = 0.87, RMSEA = 0.10, SRMR = 0.05) Strongest CFA Loading: 0.72, IRT Slope: 2.53. CNL: 4 factors -	Not reported	Not reported	Internal consistency was assessed using coefficient **Ω** for SEM and marginal reliability (θ) for IRT. Reliability was high for INL (**Ω** = 0.85, θ = 0.86) and CNL (**Ω** = 0.82–0.90). FNL showed adequate reliability (**Ω** = 0.70, θ = 0.85 for Factor 1), though Factor	Not reported	Comprehensive scope: the tool measures all three domains of NL: FNL, INL, CNL and each scale can be used separately or together. Rigorous development process involving 6 steps. Relevance to real-life contexts: includes items reflecting modern sources of nutrition information, such as social media.	Psychometric properties demonstrated only acceptable to reasonable model fit. Limited generalizability due to the predominantly White and female sample. The cross-sectional design restricts conclusions about the tool's sensitivity to change or its predictive validity regarding eating behaviors.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		acceptable model fit (CFI = 0.85, RMSEA = 0.07, SRMR = 0.06) Strongest CFA Loading: 0.92, IRT Slope: 3.29. Convergent validity: moderate correlations between nutrition literacy subdomains and fruit/vegetable intake which measured by the National Cancer Institute Fruit and Vegetable Screener (r = 0.24–0.28, p < 0.001), supporting expected relationships			2 had slightly lower consistency ($\Omega = 0.60$, $\theta = 0.76$).			

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		with dietary behavior.						
FNLI Demir & Özer (2022)	Expert-reviewed, CVI = 0.848 - 0.890. Face validity: 20 university students reviewed the survey for clarity and comprehension.	EFA (KMO = 0.606-0.821, Bartlett's Test of Sphericity $p < 0.001$, 3 domains, 36-item model, Variance Explained (Knowledge = 60%, Attitude = 55%, Behavior = 63%)), CFA ($\chi 2/df = 1.769$, GFI = 0.902, CFI = 0.895, RMSEA = 0.038, SRMR = 0.048). Discriminative validity: Women had significantly higher FNL knowledge scores than men ($p = 0.005$)	Not reported	Not reported	Cronbach's α: Knowledge α = 0.605, Attitude α = 0.761, Behavior α = 0.727	Knowledge ICC = 0.91, Attitude ICC = 0.88, Behavior ICC = 0.84	Comprehensive theoretical framework: robust framework by Vidgen & Gallegos (2014) encompassing declarative, procedural, and subjective knowledge. Rigorous development methods: a nine-step, three-phase process. Strong psychometric characteristics and good practical utility in university setting.	Limited sample scope: EFA and CFA both conducted on the same sample (n=538). Low Internal Consistency in knowledge domain: possibly due to the small number of items or their heterogeneity. Pandemic-related (COVID-19) constraints during sampling. Cultural specificity by developing in Turkish context limiting direct application.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
FLS Guine et al., (2023)	Not reported	EFA (KMO = 0.966, Bartlett's Test of Sphericity $p < 0.0005$ (significant), 3-factor model, 61.76% variance), CFA ($\chi^2/df = 3.702$, CFI = 0.904, GFI = 0.847, RMSEA = 0.076, RMR =0.054, SRMR = 0.055)	Not reported	Not reported	Cronbach's α = 0.962	Not reported	Robust psychometric validation employing EFA and CFA. High internal consistency. Conceptually well-defined dimensions: Nutritional composition, Food labelling and choices, Healthy eating practices. Large and adequate sample size (n=924)	The FLS shows promising initial structural integrity and reliability, but its overall psychometric quality is moderate at best due to absence of content, criterion, test-retest reports. The study used a convenience sample, predominantly female (79.7%), which may not be representative of the broader student populations. Recruitment methods using online tools and snowball sampling could introduce volunteer bias and limit generalizability. Cross-sectional design of study limits causal inference which

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
								hinders the results interpretation.
NLAQ Yan et al., (2023)	Not reported	EFA (KMO = 0.9, 3-factor model, 71.88% variance), CFA (CFI = 0.948, GFI = 0.929, NFI = 0.939, RMSEA = 0.082). Convergent Validity: (AVE > 0.60). Factor Loadings (λ) > 0.70 (except Q1_4), CR = 0.950. Strong but Q1_4 may need refinement. Measurement error low for most items (θ < 0.50), but Q1_4 (θ = 0.518) slightly above the threshold	Not reported	Not reported	Cronbach's α = 0.909	Not reported	Sound psychometric characteristics. Clear and comprehensive structure with three dimensions aligned with Nutbeam's (2000) model. Latent profile analysis was conducted which increase the methodological rigor of the study.	Limited Generalizability: study conducted in only two universities in Wuhan using convenience sampling. Although inspired by Nutbeam's model, the NLAQ focuses on functional literacy with limited emphasis on critical or interactive literacy. The cross-sectional design of the study did not produce highly precise results and increased uncertainty.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
S-NutLit Vrinten et al., (2023)	Expert review (n=16): S-CVI = 0.90, cognitive interviews (n=12) for face validity and readability, pre-test (n=101)	EFA: (KMO = 0.85 and 0.83, Bartlett's test of sphericity: ($\chi2$ = 1122.53, df = 66, $p < 0.001$ and $\chi2$ = 970.27 df = 55, $p < 0.001$), 2-factor model, 44.3% variance). Convergent Construct Validity: Health literacy: r = 0.27, $p < 0.001$. General literacy and numeracy: r_s = 0.12, p = 0.046. Education level: r_s = 0.13, p = 0.025. Associations with dietary behaviors and physical activity,	Not reported	Not reported	Cronbach's α = 0.80	Two-way mixed ICC = 0.74 (95% CI [0.61, 0.83], $p < 0.001$)	Rigorous methodology: the study used good practices of scale development. Assessed construct and criterion validity using established tools for health literacy, general literacy, and physical activity. Inclusive development process: the study included both experts and the target population during scale development, enhancing content relevance and clarity.	Participants were primarily from upper socioeconomic backgrounds, restricting their applicability to lower SES populations. The pre-test sample was insufficient for robust factor analysis, which could lead to excessive item reduction. Discarding elements based entirely on statistical loadings may have reduced conceptual coverage, particularly by excluding critical NL's engagement dimension. The broader topic of nutrition literacy lacks defined conceptual limits, hindering the creation of comprehensive tools.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		especially the 'Expert skills' subscale ($p < 0.001$). Unexpected association with sugar-sweetened beverage consumption.						
Iranian NL-SF12 Mostafazadeh et al., (2024)	CVI = 0.91, CVR = 0.88	EFA: (KMO = 0.840, Bartlett's Test of Sphericity: $\chi^2 = 1912.335$, $p < 0.001$, 6-factor model, 59.57% variance explained). CFA: ($\chi^2/df = 2.621$, CFI = 0.957, GFI = 0.926, TLI = 0.942, RMSEA = 0.059). Correlation with (AEBQ - Adult Eating Behavior Questionnaire).	Not reported	Not reported	Cronbach's α = 0.86	ICC = 0.83	Good alignment with the conceptual framework. Sound psychometric characteristics. Practical utility: Shortened from a 43-item tool to a concise 12-item version without compromising psychometric quality.	Limited Generalizability. self-report biases due to reliance on self-assessment. Cross-sectional design limits causal inference.

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		NL-SF12 scores and eating behaviors (Adjusted R^2 = 0.41, $p < 0.001$).						
CS-NutLit Liu et al., (2024)	Expert panel (n=6), I-CVI = 0.833 - 1.000, S-CVI/Ave = 0.908. Face validity (n=50): pre-survey assessment by young people.	EFA: (KMO = 0.857, 2-factor model, 51.029% variance), CFA: (CFI = 0.964, GFI = 0.950, TLI = 0.954, RMSEA = 0.053 and $\chi 2/df$ = 1.720). Convergent validity: AVE = 0.420 (Information), 0.515 (Expert), CR = 0.852, 0.760 Discriminant validity: HTMT = 0.515. Discriminative validity: higher literacy in those	Not reported	Translation & back-translation by Brislin's Two-Way Translation method, cultural adaptation and debugging.	Cronbach's α = 0.826	ICC = 0.818, subscales stable over two weeks (n=50).	Rigorous cross-cultural adaptation: translation followed Brislin's Two-Way Model with expert panel review, cultural adaptation, and pretesting among 50 young adults. Excellent content, construct validity, good reliability	The study focused on individual/family-level predictors such as education and parental background and did not include community or policy-level factors. Sampling bias due to convenience sampling resulted in an overrepresentation of students (54.3%), limiting representativeness across occupational groups. Samples derived from Liaoning, Shandong, and Hunan, and data collected online, which may omit

	Validity				Reliability			
Tool author (year)	**Content validity**	**Construct validity**	**Criterion validity**	**Cross-cultural validity**	**Internal consistency**	**Test-Retest reliability**	**Strengths**	**Limitations**
		with higher education (p = 0.002), nutrition courses (p < 0.001), and more exposure to health info (p < 0.001).						offline or rural populations.

AVE, Average Variance Extracted; CFA; Confirmatory Factor Analysis; CR, Composite Reliability; CVI, Content Validity Index; I-CVI, Item-level Content Validity Index; S-CVI, Scale-level Content Validity Index; CVR, Content Validity Ratio; CCR, Composite Construct Reliability; IRT, Item Response Theory; NL-TCS, Nutrition Literacy for Taiwanese College students; Turkish SFLQ, Turkish Short Food Literacy Questionnaire; NLI-28, Nutrition Literacy Inventory – 28 items; CNLT-R, Critical Nutrition Literacy Tool – Revised; KFLS, Korean Food Literacy Scale; EFLBQ, Eating and Food Literacy Behaviors Questionnaire; NL-SF12, Nutrition Literacy – Short Form 12 items; Spanish SPFL, Spanish Self-perceived Food Literacy; Korean SPFL, Korean Self-perceived Food Literacy; YA-NLT, Young Adults – Nutrition Literacy Tool; FNLI, Food and Nutrition Literacy Instrument; FLS, Food Literacy Scale; NLAQ, Nutrition Literacy Assessment Questionnaire; S-NutLit, Short-Nutrition Literacy; Iranian NL-SF12, Iranian Nutrition Literacy – Short Form 12 items; CS-NutLit, Chinese Short – Nutrition Literacy; PCA, Principal Component Analysis; SEM, Structural Equation Model; EFA, Explanatory Factor Analyses; KMO, Kaiser-Meyer-Olkin; α, Cronbach's alpha; Ω, McDonald's omega; r_s, Spearman's correlation coefficient; ICC, IntraClass Coefficient; GFI, Goodness-of-Fit Index; AGFI, Adjusted Goodness-of-Fit Index; RMSEA, Root mean square error of approximation; CFI, Comparative Fit Index; TLI, Tucker-Lewis index; RMR, Root Mean Square Residual; SRMR, Standardized Root Mean Square Error of Approximation; NFI, Normed Fit Index; HTMT, Heterotrait-Monotrait Ratio of Correlations.

4.6 Quality appraisal of tools

The majoirty of FL and NL tools demonstrated strong internal consistency, with most tools having Cronbach's α values more than 0.80, indicating good internal structures. Exploratory and confirmatory factor analyses were used consistently and produced good to excellent fit indices (CFI ≥ 0.90, RMSEA < 0.08), demonstrating construct validity for most instruments. However, while content validity procedures were extensively reported and rigorously analyzed by expert panel reviews and Delphi studies, criterion validity, test-retest reliability, and hypothesis testing were infrequently reported, leaving significant gaps for future validation studies. The COSMIN risk of bias evaluation highlights these gaps, as some psychometric domains were not consistently reported implying the need for more extensive methodological reporting. The detailed report of COSMIN assesment can be find in Appendices section (Supplementary table 1).

The psychometric properties of the FL and NL tools assessed in this review had different levels of validity and reliability, indicating the presence of strengths and limitations in the development process. Expert reviews or Delphi panels supported the content validity of most tools, with high CVI scores recorded, in NL-TCS (CVI = 1.00), NLI-28 (CVI ≥ 0.90), and CS-NutLit (S-CVI/Ave = 0.908). EFA and CFA were often used to examine construct validity, for example, KFLS indicated strong construct validity (8-factor structure, RMSEA = 0.064), whilst NL-SF12 and S-NutLit provided good model fits. Criterion validity was found less frequently due to absence of ‘gold standard’ or widely accepted objective outcome. Cross-cultural adaptation was addressed in translations such as the Spanish SPFL and CS-NutLit by using back-translation and cultural adaptation, but statistical testing inlcuding measurement invariance was not conducted indicating the less rigor. Cronbach's α indicated acceptable to high internal consistency across tools (Cronbach's α = 0.70-0.90). However, some domains in FNLI (e.g., Knowledge α = 0.605) and CNLT-R (α = 0.69) had lower

borderline values. NLAQ and FLS have less methodological rigor in testing content, criterion, cross-cultural validities as well as test-retest reliability. When test-retest reliability was reported (e.g., Turkish SFLQ r_s = 0.808, KFLS: ICC = 0.90), it indicated temporal stability. Despite largely positive psychometric profiles, limitations included limited generalizability due to convenience sampling in the majority of tools, a lack of external validation (e.g., CNLT-R, FLS), and no examination of responsiveness to changes in FL/NL over time.

Nutrition Literacy Tool for Taiwanese College Students (NL-TCS; Liao et al., 2018)

The NL-TCS stands out by its dual assessment approach which includes both self-perceived and scenario-based measures of NL. The development process of the tool was evaluated as adequate due to absence of qualitative data collection such as cognitive interviews about the comprehensibility of the tool from students not experts. It has great psychometric integrity, with excellent internal consistency (Cronbach's α = 0.81) for scenario-based part and (Cronbach's α = 0.85) for self-rated part, also with very good construct and convergent validity according to COSMIN grading. Construct/Structural validity confirmed by CFA and convergent validity confirmed statistically significant correlation with both the self-rated and the scenario-based scales positively correlated with healthy-eating behavior scale (r = .417, p < .001 and r = .125, p < .001, respectively). The structural validity analyses produced remarkable fit indices for self-rated part: GFI = 0.993 > 0.9, AGFI = 0.988 > 0.9, RMSEA = 0.019 < 0.08; for scenario-based part: GFI = .926 > 0.9, AGFI = .913 > 0.9, RMSEA = 0.015 < .08. The tool demonstrated statistical rigor by adopting IRT approaches with Rasch model for item-level difficulty analysis along with verification of unidimensionality through PCA. Limitations included a strong ceiling effect with 77.4% of correct answers in scenario-based part of the tool indicating items with

insufficient difficulty difference. Additionally, the use of convenience sampling techniques can limit the generalizability of results. The tool is commonly cited and used in East Asian research contexts with 13.89% of overall usage after search engine results (see Supplementary table 2) and is suggested for localized assessments.

Turkish Short Food Literacy Questionnaire (SFLQ; Durmus et al., 2019)

The Turkish SFLQ is psychometrically strong scale with expert consensus and excellent internal consistency ($\alpha = 0.803$). The structural validity study revealed satisfactory construct representation, while explaining a lower proportion of variation (32.01%) than the original scale (76.4%) (Gréa Krause et al., 2018). The Turkish version of the SFLQ tested construct validity via hypothesis testing. The study hypothesized and confirmed that higher health literacy scores, measured by TSOY-32 and NVS would be positively associated with SFLQ scores. Significant correlations ($r = 0.531$ and $r = 0.294$, $p < .001$) support convergent validity and the comparator tools used are psychometrically robust. The COSMIN assessment of the Turkish SFLQ shows “very good” scores for criterion validity, and internal consistency with hypothesis testing; “adequate” score for structural validity due to absence of CFA; and “doubtful” score for test-retest reliability due to absence of ICC testing. The original scale on which it is based is the most frequently used tool among all those reviewed, accounting for 27.78% of usage across reviewed studies, reflecting its wide acceptance and applicability in research and practice. Turkish version of SFLQ excel with cultural adaptation through rigorous translation, practical utility as a brief 12-item tool, and cross-cultural consistency by replicating the original scale's unidimensional structure and reliability ($\alpha=0.82$). Nonetheless, certain methodological constraints, such as social desirability biases inherent in the self-report design and lower variance compared to original scale can be a potential limitations.

Nutrition Literacy Inventory (NLI-28; Makiabadi et al., 2019)

The NLI-28 demonstrated excelent construct validity with a 4-factor model accounting for 78% of the variation and good CFA indices (e.g., RMSEA = 0.002, CFI = 0.963, GFI = 0.942, AGFI = 0.910, IFI = 0.957). The tool's high internal consistency (α = 0.87) and cross-cultural adaption processes shows its psychometric strengths. Construct validity was well reported, with significant convergent validity (r = 0.74) and divergent validity (r = - 0.11) relationships. According to the COSMIN assessment, the scale received "very good" scores for structural validity and internal consistency, but inadequate scores for criterion and convergent validities though the correlation coefficients were given, the correlation with what kind of outcomes or measures is not specified explicitly hindering the overall understanding. The known-groups validity was evaluated as "doubful" due to poor description of the important characteristics of the subgroups. The frequency of NLI-28 use compared to other tools remains low (2.78%), with only one study citing it as a result of search. Limitations involved the use of cluster random sampling within one specific university which can reduce generalizability, and the self-administered design which presents the possibility of response bias.

Critical Nutrition Literacy Tool–Revised (CNLT-R; McNamara et al., 2020)

The CNLT-R demonstrated moderate psychometric characteristics due to absence of majority of domains according to COSMIN risk of bias checklist and borderline values of internal consistency, however it has shown statistical rigor in contruct/structural validity testing by using both CTT and IRT approaches in development stages. CNLT-R is one of the tool which emphasized the importance of critical media literacy and evidence-based nutrition sources appraisal among young adults which is actual in university settings where social media is a substantial part of life. Despite a strong psychometric foundation supported by

EFA and CFA outcomes (e.g., CFI = 0.90, RMSEA = 0.11, SRMR = 0.07), its internal consistency was slighlty low nearly meeting borderline value between good and doubtful values (Cronbach's α = 0.69), particularly for its second factor named ‘critical appraisal of evidence-based nutrition sources’ indicating potential conceptual ambiguity. The demographic data which largely consists of white and female can considerably limit its applicability to increasingly varied groups. The CNLT-R, which has the potential for wider application, would benefit from additional validation studies to improve demographic inclusivity and internal consistency. According to COSMIN evaluation, structural validity of CNLT-R is very good, but the internal consistency is doubtful due to borderline value. The frequency of tool usage is low (2.78%). But for now, this may be due to the fact that this is a revised version of the Norwegian version tailored to the specifics of U.S college settings.

Korean Food Literacy Scale (KFLS; Na & Cho, 2020)

The KFLS developed by Na & Cho (2020) is a psychometrically strong and culturally grounded tool with ‘very good’ COSMIN grades in development, structural validity, internal consistency, and reliability characteristics. Tool has undergone rigorous three-phase development procedures involving Delphi study with experts, a pilot test, and a confirmatory study. Despite its lack of wide usage, it provides a robust 8-factor model with CFA supporting the structural validity (CFI = 0.976, TLI = 0.975, RMSEA = 0.064), high internal consistency (Cronbach’s α = 0.737–0.910), and test-retest reliability (ICC = 0.90). The tool covered cultural specificity, comprehensive content covering multiple aspects of FL, and methodological rigor. Limitations such as restricted regional sampling lowering national representativeness, lengthy content consisting of 50 items, and absence of benchmark against other related instruments or correlation with dietary intake measures or health outcomes limiting its external validity.

Eating and Food Literacy Behaviors Questionnaire (EFLBQ; Rhea et al., 2020)

The EFLBQ has demonstrated a strong psychometric foundation through extensive item generation, exploratory and confirmatory factor analysis, and test-retest reliability which confirmed by 'very good' COSMIN scores for structural validity, internal consistency. However the tool is assessed as 'doubtful' for test-retest reliability domain due to absence of ICC testing which is according to COSMIN more rigorous than Pearson correlation. The tool has moderate usage frequency (11.11%), indicating that it is increasingly being used in research. Its characteristics include a robust 5-factor model (KMO = 0.817; CFA: CFI = 0.98, TLI = 0.98, RMSEA = 0.06, SRMR = 0.05). The tool has a good internal consistency range of 0.63-0.89. The tool covers highly relevant to students behavioral domains such as 'health and nutrition', 'convenience', and 'taste' along with Vidgen & Gallegos (2014) 'food preparation', 'planning and decision-making' which increases its real-world applicability. Some limitations exist such as convenience sampling technique mostly including White and female demographics, potential education as the findings may not apply to individuals with a lower educational level, and slighlty low internal consistency in some subscales. Overall, the EFLBQ is recommended for use in a variety of settings, especially in higher education and intervention studies where practical food behaviors are important.

Nutrition Literacy Short Form-12 (NL-SF12; Mo et al., 2022)

The NL-SF12, a brief form of original NL-43 tool, psychometrically validated assessment tool with very good COSMIN ratings for structural validity and internal consistency as well as criterion and convergent validities due to good correlation with original NL-43: r = 0.969. Despite its comparatively low usage frequency (2.78%) in other studies, the CFA shows great construct validity (CFI = 0.972, GFI = 0.960, AGFI = 0.919, RMSEA = 0.069) and high internal consistency (Cronbach's α = 0.890). NL-SF12 was

developed for Chinese college students, and exceled with cultural relevance, efficient and practical short-form administration with strong psychometric validation, and great alignment with the original scale. Limitations of the tool include limited generalizability due to validation in a single area sample and lower convergent and discriminant validities in the "Obtain" and "Apply" subscales. The study did not conduct hypothesis testing to test convergent validity against external variables which can limit its validation scope and there is a risk of compomising valifity and reliability as a result of shortening the measuring scale.

Spanish Self-Perceived Food Literacy (SPFL; Luque et al., 2022)

The Spanish SPFL scale is a psychometrically sound tool with very good COSMIN ratings for structural validity and internal consistency, and a high usage frequency of original scale from which it is derived (25%), indicating high utility in the field. The tool has a well-supported 5-factor structure with good fit (CFA: CFI = 0.95, TLI = 0.94, RMSEA = 0.045, GFI = 0.91) and high internal consistency (Cronbach's $\alpha = 0.894$). It demonstrated a meaningful positive correlations with Mediterranean diet adherence ($r = 0.59$) and negative correlation with psychological traits such as impulsivity ($r = -0.26$) indicating the tool's good convergent and discriminant validity. The limitations emphasized by authors of tool include the representation of only four domains of Vidgen and Gallegos' (2014) framework but not all eleven conceptual components that potentially limits the breadth of its construct coverage. Also, the relatively limited demographic with notable underrepresentation of male students, further restricts generalizability.

Korean Self-Perceived Food Literacy (Korean SPFL; Lee et al., 2022)

The Korean SPFL scale is a validated tool with very good COSMIN scores for structural validity, internal consistency, convergent, and discriminative validity. The original

scale (SPFL) usage frequency of the tool is high (25%). It is based on a 6-factor model with a strong CFA finding (NFI = 0.860, CFI = 0.90, TLI = 0.89, RMSEA = 0.08), and demonstrates high internal consistency range (Cronbach's α = 0.840–0.929). Major strengths of the tool include the incorporation of concepts associated with sustainability and validated correlations with dietary behaviors using SEM analysis. The limitations are its self-report administration, sampling restricted to Seoul-based students, and limited generalizability. The tool is recommended for Korean university settings, especially in sustainability-oriented food literacy studies.

Young Adults Nutrition Literacy Tool (YA-NLT; McNamara, 2022)

The YA-NLT displayed high methodological rigor through a six-step process that included focus groups (n = 24), expert panel review (n = 10), iterative item refinement, and psychometric testing based on both Classical Test Theory (CTT) and Item Response Theory. It evaluates all three domains of NL, expanding conceptual coverage and providing the option to use each subscale separately. The tool has "very good" COSMIN scores for structural validity, internal consistency, and for convergent validity. According to CFA analysis FNL subscale had excellent model fit (CFI = 0.99, RMSEA = 0.04) and CNL and INL subscales had acceptable-to-moderate fit (CFI = 0.85-0.87) with IRT slopes confirming good item discrimination across domains. The tool has a strong theoretical foundation which is relevant to modern context where digital and social media plays an important role. Using several validation procedures, such as IRT and CFA, and reliability indicators (Ω = 0.70-0.90), enhances the psychometric credibility. Limitations involved a lack of test-retest reliability and low demographic diversity with the majority of participants being White and female which can limit generalizability. The tool's responsiveness and predictive validity can be compromised due to the cross-sectional design. Despite its present utilization rate of being

low, the YA-NLT is highly recommended for future research due to its conceptual comprehensiveness, psychometric strength, and methodological rigor.

Food and Nutrition Literacy Instrument (FNLI; Demir & Özer, 2022)

The FNLI is a methodologically robust tool developed through a nine-step, three-phase procedure based on Vidgen and Gallegos' (2014) framework. The scale evaluates knowledge, attitudes, and behaviors across three domains with great structural validity (CFA: RMSEA = 0.038, CFI = 0.895). The attitude and behavior subscales have high internal consistency (Cronbach's $\alpha \geq 0.72$), but the knowledge subscale has lower consistency ($\alpha = 0.605$). Test-retest reliability value was excellent, with ICC values ranging from 0.84-0.91. The tool's psychometric soundness is supported by very good COSMIN scores for structural validity, internal consistency, reliability, and hypothesis testing of known-groups validity which confirmed the gender-based difference in literacy scores. Strengths of the tool included a solid theoretical foundation and domain-specific structure. However, its generalizability is limited by low reliability in the knowledge area, the use of a single sample for EFA and CFA, and the low frequency of application. FNLI is suggested for both local use and future cross-cultural validation.

Food Literacy Scale (FLS; Guine et al., 2023)

The FLS is a well-structured scale with promising psychometric characteristics in the field. The tool is based on three factors: nutritional content, food labeling, and healthy eating practices, with high internal consistency (Cronbach's $\alpha = 0.962$). Its structural validity is supported by strong CFA results (CFI = 0.904, RMSEA = 0.076), and it has achieved very good COSMIN scores in these areas. The tool, which was developed with a large sample size ($n = 924$), has the potential for widespread use. But FLS has some limitations such as the lack

of data on content, convergent, dicriminative validities, and test-retest reliability of the tool which are required for a rigorous psychometric evaluation. Also, the study used convenience sampling with a mostly female sample, and its cross-sectional design limits its capacity to examine changes over time. With a slightly low usage frequency (5.56%), the FLS is still requires additional validation.

Nutrition Literacy Assessment Questionnaire (NLAQ; Yan et al., 2023)

The NLAQ is a concise and psychometrically sound tool which demonstrated strong structural validity with a 3-factor model (CFA: CFI = 0.948, GFI = 0.929, NFI = 0.939, RMSEA = 0.082) and high internal consistency (Cronbach's $\alpha = 0.909$). The tool received 'very good' COSMIN ratings for structural validity and internal consistency and discriminative or known-groups validity. Developed with methodological rigor, including latent profile analysis to identify literacy subgroups, the NLAQ provides clear, well-defined constructs. However, the tool has a very low usage frequency (2.78%), and limitations include reliance on convenience sampling from only two universities in Wuhan, reducing generalizability and narrowly focused on functional literacy omitting interactive and critical domains. The NLAQ is one of the strong candidate for functional nutrition literacy assessment in academic settings but further broader application and expansion to cover more comprehensive literacy domains is required.

Short-Nutrition Literacy (S-NutLit; Vrinten et al., 2023)

The S-NutLit is one of the brief yet methodologically sound tools which was developed specifically for young adults. It demonstrated very good ratings for internal consistency, convergent, discriminative validities, and reliability, adequate score for structural validity due to absence of CFA according to COSMIN, and has a moderate usage frequency

(5.56%). Developed through expert review, cognitive interviews, and pre-testing, it reflects strong methodological rigor. Psychometrically, the tool is developed based on a 2-factor structure with acceptable variance explained (44.3%) and showed good internal consistency (Cronbach's α = 0.80), and test-retest reliability (ICC = 0.74). Convergent validity was established by correlations with health literacy (r = 0.27, $p < .001$), education level, and dietary behaviors in the "Expert skills" subscale. Strengths included an inclusive development process involving both experts and target population, clear subscale structure, and strong methodology. Limitations included overrepresentation of participants from higher socioeconomic backgrounds, a pre-test sample too small for factor analysis, and potential conceptual narrowing due to item reduction.

Iranian Nutrition Literacy Short Form-12 (Iranian NL-SF12; Mostafazadeh et al., 2024)

The Iranian NL-SF12 is a culturally adapted, psychometrically robust short-form tool designed to assess nutrition literacy among university students in Iran. It achieved very good COSMIN ratings for structural, convergent, discriminative validities, internal consistency, and reliability. Developed through a rigorous translation and validation process, the scale maintained a strong 6-factor structure (CFA: CFI = 0.957, GFI = 0.926, TLI = 0.942, RMSEA = 0.059) and showed high internal consistency (α = 0.86). Limitations included cross-sectional design that can limit causal inferences between nutrition literacy and behavioral outcomes, study was limited to a single regional sample with a relatively small and localized sample, use of self-report questionnaires introduces the possibility of response bias.

Chinese Short-Nutrition Literacy (CS-NutLit; Liu et al., 2024)

The CS-NutLit is a newly developed, culturally adapted instrument designed to assess nutrition literacy among young Chinese adults, particularly university students. It received

very good COSMIN ratings for structural, discriminative validity, internal consistency, and reliability. Developed through a three-phase process, including Brislin's two-way translation, cognitive pre-testing, and expert panel validation, the CS-NutLit demonstrated a robust 2-factor structure (Information and Expert Skills) with strong model fit (CFI = 0.964, GFI = 0.950, TLI = 0.954, RMSEA = 0.053) and high internal consistency (Cronbach's α = 0.826) and test-retest reliability (ICC = 0.818). Strengths include its solid theoretical foundation, sound methodological design, and successful cross-cultural adaptation, ensuring high relevance for Chinese-speaking student populations. While the tool effectively assess individual-level determinants of NL, it does not include community or policy-level influences, limiting its scope for broader public health applications and samples derived online, which may omit offline or rural populations. Nonetheless, the CS-NutLit is highly suitable for use in NL research and intervention design within university contexts in China.

4.7 Conclusion

The 16 tools were identified as a result of a systematic review with differing concepts on which they are based, the number of items and constructs, and psychometric properties. The main general characteristics, such as tool name, authors, country of origin, measured concept (FL and/or NL), purpose, conceptual framework, number of items and constructs, development methods, scoring details, and sample characteristics were summarized in Table 1. The psychometric characteristics were summarized in Table 2, where validity and reliability, as well as strengths and limitations of tools, were outlined. All these properties from two tables were narratively synthesized, and then for each tool was given a recommendation where strengths and limitations (Table 2), COSMIN score (Supplementary table 1), and approximate frequency of usage (Supplementarty table 2), which was identified through the search in the same academic databases, were reported.

CHAPTER V: DISCUSSION

The systematic review synthesized and critically evaluated 16 validated tools designed to measure FL and NL among university students' population. The analysis aimed to analyze how FL and NL are conceptualized, how they are measured, and the psychometric properties of tools.

The review revealed substantial heterogeneity in the conceptualization of FL and NL, though the convergence exists regarding the underlying theoretical frameworks such as Nutbeam's tripartite health literacy model and Vidgen & Gallegos's framework. FL is generally conceptualized as a multidimensional construct encompassing knowledge, skills, and behaviors that enable individuals to plan, manage, select, prepare, and consume food in a health-promoting way. FL instruments, including the EFLBQ, Turkish SFLQ, Spanish and Korean SPFL, KFLS, and FNLI, were based, explicitly or implicitly, on the framework proposed by Vidgen & Gallegos (2014). FL is typically defined through dimensions involving practical food-related competencies such as planning and decision making, food selection, preparation, and budgeting. While NL tools in contrast, are typically conceptualized through Nutbeam's tripartite model of health literacy, which comprises functional, interactive, and critical domains. NL tools operationalized NL as the capacity to obtain, process, understand, and critically evaluate nutrition information for informed decision-making. Some instruments, like the S-NutLit, CNLT-R and YA-NLT, extended the conceptual scope by incorporating digital literacy and the critical appraisal of nutrition misinformation marking a shift towards more context-aware understandings of NL, which applies to university students' life experiences. Though the wordings of the domains of some NL tools can slightly differ from each other, they still cover the main three health literacy domains.

The overall quality of measurement tools can be assessed as having acceptable to strong psychometric properties, with all tools demonstrating excellent structural validity and

internal consistency, which were also supported by the COSMIN risk of bias checklist. However, not all tools underwent test-retest reliability assessment, criterion validation, and hypothesis testing. No one tool was assessed for cross-cultural validation, measurement error, and responsiveness over time. Despite the presence of culturally adapted tools that underwent expert review and rigorous translation procedures, statistical validation methods such as measurement invariance and multi-group confirmatory factor analyses were not conducted. The current literacy tools are not aimed at exploring longitudinal effects of literacy on the dietary outcomes, which is also supported by the systematic review of FL and NL tools among children and adolescents (Carroll et al., 2022). This aspect of scale development methodologies can be one of the future directions that can be explored more widely.

According to the COSMIN risk of bias checklist by Mokkink et al. (2018), the 'gold standard' in criterion validity evaluation should be decided by the review team. As is known in the field of FL and NL measurement tools, there is no definite 'gold standard' measure, and it is highly subjective to state that the field should be compared to one particular tool. According to widely used definitions of criterion validity, the measures have good criterion validity if they correlate well with other measures measuring the same construct or with 'gold standard' established in the field. Therefore, it was decided to evaluate as "Very Good" the criterion validity box in case the tool is a revised, modified, or shortened version of the original scale or with another tool measuring exactly FL or NL or together, and it was explicitly reported the correlation between them. Otherwise, the box was assessed as 'Not Reported' even though the authors of study explicitly indicated the corresponding sections as criterion validation.

Findings align well with established conceptual frameworks distinguishing between FL as more broader and practical food competencies and NL as deeper nutritional cognitive abilities encompassing critical evaluation of nutrition knowledge. However, it can be

observed that the divergence in the newer instruments lies in the broadening of the FL construct to include individual, psychosocial, and environmental determinants, particularly those that are relevant to young adults in university settings. Deliens et al., (2014) determinants of eating behavior supports EFLBQ's new domains of FL inlcuding health and nutrition, taste, and convenience among university students along with Vidgen & Gallegos food preparation, and planning, and decision-making domains. EFLBQ did not align completely with Vidgen & Gallegos (2014) framework but rather complemented it by additional personal factors. Previous studies conducted with young adults have shown that "taste", "nutrition", "cost", and "convenience" are important determinants of young adults' food choices (Aggarwal et al., 2016). Taste is considered as the most important factor in food selection among young adults (Deshpande et al., 2009; Hebden et al., 2015). That is why the inclusion of these personal factors is well-justifed during conceptualiztion and measurement of FL among university students. The original scale, SPFL (Poelman et al., 2018), based on which two tools were developed (Spanish and Korean SPFL) also shows a similar scope to EFLBQ in terms that both of them complemented Vidgen & Gallegos framework by personal factors; SPFL similar to EFLBQ examined "resilience and resistance" and "social and conscious eating", when attempting to conceptualize FL. Korean SPFL in turn, broadened the concept of FL further into sustainable eating behaviors and confirmed that "healthy snack habits", "intrinsic motivations such as self-control and self-regulation", and "knowledge about the origin of food and label information" had the significant effect on increasing the ecological eating behavior (Y. Lee et al., 2022). Also Korean SPFL noted the irrelevance of food preparation skills and social and conscious eating to sustainable eating by denoting the lack of cooking experience and less time to social eating due to technological progress along with socioeconomic and cultural transformations. The Spanish SPFL (Luque et al., 2022) excluded the factor associated with social eating due to its irrelevance to university students

who tend to give priority to their studies rather than healthy eating in a social way. However, Luque et al. (2022) stated that “there is no empirical evidence on the life experience of the target population; instead, they relied on experts for the development of the scale or followed nutritional recommendations from the country”. This can underestimate the influence of geography and culture on the concept and demonstrates psychometric differences between scales across countries. This underlines the importance of cultural adaptation and validation in scale development. Spanish SPFL differed from the original scale by keeping 5-factor model which included: “cooking skills”, “emotional management”, “healthy consumption”, “nutritional literacy & planning”, and “availability of UPF”. This showed not a full conceptual coverage of FL according to Vidgen & Gallegos (2014) due to the presence of emotional management and availability of UPF factors. KFLS covered two domains of FL: Knowledge and Skills/Ability. The tool was supported by Poelman (2018) and Krause (2016). The tool is not explicilty based on Vidgen & Gallegos (2014) framework but rather on knowledge and skills of the individuals emphasizing that increases in cooking skills, knowledge, and self-efficacy influence food choices. It is also notable the presence of broader domains of FL such as food safety, food system, food resource management, personal factor like self-efficacy in KFLS (Na & Cho, 2020). The FLS by Guine (2023) emphasized the importance of food labeling as the students tend to not pay attention to that aspect of food choice and also the lower adherence to Mediterranian diet among students implying the lower FL which was also supported by Spanish SPFL developed by Luque et al. (2022). FLS out of other FL tools also do not follow particular framework or definition.

In NL measurement, NL-TCS showed that although two parts of the NL-TCS: self-rated part and scenario-based part were significantly related; the correlation strength was lower than expected. Therefore, it was suggested by Liao et al. (2018) to further explore these two parts of the measurement as there is indeed a gap between individuals’ subjective

perceptions and their actual ability relating to NL. It was recommended that cross-referencing between both parts should be conducted and future research should explore these related but separate components of NL in detail. A notable theoretical advancement appears in tools like the CNLT-R, S-NutLit, and YA-NLT, which extend the critical literacy dimension to include digital nutrition information literacy, resistance to misinformation, and media skepticism, constructs that were largely absent in earlier NL tools reviewed by Yuen et al. (2018). These additions reflect a contemporary response to the proliferation of online dietary misinformation and the need for students to critically appraise digital health content.

An insight that diverges from earlier reviews is the lack of tools explicitly tailored to the university student population, despite this group's unique dietary challenges, including limited time, financial constraints, and increased autonomy. Previous reviews have tended to focus broadly on adults or adolescents without sufficiently accounting for the transitional context of university life. Tools such as EFLBQ, YA-NLT, Korean SPFL, and KFLS help to fill this gap by directly addressing factors such as budgeting, cooking self-efficacy, and behavioral features, all highly relevant to students navigating in food systems independently for the first time. The findings of this review suggest that while many instruments align well with foundational FL and NL frameworks, the field is evolving to accommodate new determinants (ecological, digital, personal factors and motivations), new formats (short-form, behavior-based tools), and new populations (university students). These shifts reflect an evolving field, moving from static, knowledge-focused assessments to dynamic, context-sensitive tools capable of informing real-world interventions and public health policies regarding university students. Furthermore, current review supports Yuen et al. (2018) finding that earlier tools focused mainly on functional literacy and this review showed further advancements in the field with recent tools such as CNLT-R, S-NutLit, YA-NLT that incorporate critical and digital literacy skills. While compared to Carroll et al. (2022), the

review confirms that university students are cognitively ready to engage with all FL/NL domains, bridging the gap between child and adult-focused tools through age-specific developmentally appropriate scales. Regarding the alignment of tools with Vidgen & Gallegos (2014) framework as stated by Amouzandeh (2019), not all tools aligned exactly well with the framework. Reviewed tools were mostly informed by the framework with some modifications after factor analysis.

The systematic review consolidates FL and NL scales validated explicitly among university populations. By explicitly differentiating between FL and NL constructs and comparing instruments across diverse groups of university students, this synthesis strived to clarify conceptual boundaries and overlaps that can enrich theoretical understanding. Additionally, the structured comparison of psychometric quality across reviewed instruments contributes methodologically that can guide researchers toward selecting reliable tools for health promotion and educational interventions. According to the results of review, there is a clear necessity for developing or adapting measurement tools specifically related to the unique nutritional and food literacy needs of university students. The great theoretical variations in defining FL and NL offered the great variety of tools which complicates the development of universal university based assesment tools. The current review demonstrates a gap in tools that comprehensively integrate both FL and NL dimensions relevant to young adults transitioning toward independent food-related decision-making. Particularly, longitudinal research to examine causal relationships between literacy interventions and dietary outcomes, cross-cultural validation across tools, and adoption of consensus-based methodology based on universal conceptual definition is critically needed.

This comprehensive synthesis clearly separates and integrates FL and NL as distinct but overlapping constructs crucial for understanding food patterns among university students. The validated tools provide high-quality assessments, making substantial contributions to

educational practice and public health research. This synthesis, by identifying specific implications for future research, practice, and policy, lays the theoretical groundwork for developing targeted interventions to improve university students' food-related competencies and skills.

Limitations

The systematic review has several limitations. First, although a comprehensive search and inclusion strategy was established, the systematic review was limited to studies published in English and available in full-text format which may have excluded relevant tools developed in non-English-speaking contexts. Second, while the COSMIN checklist was applied, some studies lacked reporting on key psychometric properties such as test–retest reliability, responsiveness, and measurement error, which constrained full quality appraisal. Third, despite efforts to capture cultural adaptations, the level of cultural contextualization was frequently inconsistently reported across reviewed studies, making it difficult to make clear conclusions about cross-cultural equivalence.

CHAPTER VI: CONCLUSION

In conclusion, this systematic review presents a targeted and evidence-based synthesis of FL and NL measurement tools validated for university students, a demographic that has been frequently overlooked in previous reviews. The findings show that, while foundational frameworks like Nutbeam's health literacy model and Vidgen and Gallegos' food literacy model continue to be influential, recent instruments have improved in terms of conceptual depth and psychometric sophistication. Tools such as EFLBQ, CNLT-R, YA-NLT, and KFLS indicate a significant move toward multidimensional evaluations that cover critical thinking, digital literacy, sustainability, and behavioral self-regulation—elements applicable to the university context. Moreover, the application of the COSMIN risk of bias checklist enabled an evaluation of tool quality, revealing notable strengths in internal consistency and structural validity, while also identifying gaps in test–retest reliability, cross-cultural adaptability, and responsiveness. This systematic review by bridging the gap between child- and adult-oriented tools in the field of nutrtition and emphasizing the importance of contextually and culturally driven tools for university students lays the groundwork for future research and educational interventions aimed at improving food and nutrition literacy among young adults.

References

Aggarwal, A., Rehm, C. D., Monsivais, P., & Drewnowski, A. (2016). Importance of taste, nutrition, cost and convenience in relation to diet quality: Evidence of nutrition resilience among US adults using National Health and Nutrition Examination Survey (NHANES) 2007–2010. *Preventive Medicine*, *90*(2016), 184–192. https://doi.org/10.1016/j.ypmed.2016.06.030

Aihara, Y., & Minai, J. (2011). Barriers and catalysts of nutrition literacy among elderly Japanese people. *Health Promotion International*, *26*(4), 421–431. https://doi.org/10.1093/heapro/dar005

Alkazemi, D. (2019). Gender differences in weight status, dietary habits, and health attitudes among college students in Kuwait: A cross-sectional study. *Nutrition and Health*, *25*(2), 75–84. https://doi.org/10.1177/0260106018817410

Amin, S. A., Lehnerd, M., Cash, S. B., Economos, C. D., & Sacheck, J. M. (2019). Development of a Tool for Food Literacy Assessment in Children (TFLAC). *Journal of Nutrition Education and Behavior*, *51*(3), 364–369. https://doi.org/10.1016/j.jneb.2018.12.006

Amouzandeh, C., Fingland, D., & Vidgen, H. A. (2019). A scoping review of the validity, reliability and conceptual alignment of food literacy measures for adults. *Nutrients*, *11*(4). https://doi.org/10.3390/nu11040801

Ashoori, M., Omidvar, N., Eini-Zinab, H., Shakibazadeh, E., & Doustmohamadian, A. (2020). Development and Validation of Food and Nutrition Literacy Assessment Tool for Iranian High-school Graduates and Youth. *International Journal of Preventive Medicine*, *8*, 1–14. https://doi.org/10.4103/ijpvm.IJPVM

Barbour, L. R., Ho, M. Y. L., Davidson, Z. E., & Palermo, C. E. (2016). Challenges and opportunities for measuring the impact of a nutrition programme amongst young people

at risk of food insecurity: A pilot study. *Nutrition Bulletin*, *41*(2), 122–129. https://doi.org/10.1111/nbu.12200

Bari, N. N. (2012). *MASTER THESIS Master ' s Programme in Food , Nutrition and Health NUTRITION LITERACY STATUS OF ADOLESCENT STUDENTS IN KAMPALA DISTRICT , UGANDA*. *January 2012*, 1–153.

Begley, A., Paynter, E., & Dhaliwal, S. S. (2018). Evaluation tool development for food literacy programs. *Nutrients*, *10*(11). https://doi.org/10.3390/nu10111617

Boucher, B. A., Manafò, E., Boddy, M. R., Roblin, L., & Truscott, R. (2017). The Ontario food and nutrition strategy: Identifying indicators of food access and food literacy for early monitoring of the food environment. *Health Promotion and Chronic Disease Prevention in Canada*, *37*(9), 313–319. https://doi.org/10.24095/hpcdp.37.9.06

Carroll, N., Perreault, M., Ma, D. W. L. L., & Haines, J. (2022). Assessing food and nutrition literacy in children and adolescents: a systematic review of existing tools. *Public Health Nutrition*, *25*(4), 850–865. https://doi.org/https://doi.org/10.1017/S1368980021004389

Colatruglio, S., & Slater, J. (2014). Food Literacy: Bridging the Gap between. *Sustainable Well-Being: Concepts, Issues, and Educational Practices*, *December*, 37–55.

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, *22*(10), 1435–1443. https://doi.org/10.1177/1049732312452938

Cullen, T., Hatch, J., Martin, W., Higgins, J. W., Sheppard, R., Cullen RD, MS, T., Hatch RD, MHSc, J., Martin RN, PhD, W., Higgins PhD, J. W., & Sheppard RN, BScN, R. (2015). Food literacy: Definition and framework for action. *Canadian Journal of Dietetic Practice and Research*, *76*(3), 140–145. https://doi.org/10.3148/cjdpr-2015-010

Deesamer, S., Piaseu, N., Maneesriwongul, W., Orathai, P., & Schepp, K. G. (2020). Development and psychometric testing of the Thai-nutrition literacy assessment tool for

adolescents. *Pacific Rim International Journal of Nursing Research*, *24*(1), 5–19.

Deliens, T., Clarys, P., De Bourdeaudhuij, I., & Deforche, B. (2014). Determinants of eating behaviour in university students: a qualitative stud...: Discovery Service for Endeavour College of Natural Health Library. *BMC Public Health*, *14*(53), 1–22. http://www.biomedcentral.com/1471-2458/14/53%0ARESEARCH

Demir, G., & Özer, A. (2022). Development and validation of food and nutrition literacy instrument in young people, Turkey. *Progress in Nutrition*, *24*(2). https://doi.org/10.23751/pn.v24i4.13051

Deshpande, S., Basil, M. D., & Basil, D. Z. (2009). Factors influencing healthy eating habits among college students: An application of the health belief model. *Health Marketing Quarterly*, *26*(2), 145–164. https://doi.org/10.1080/07359680802619834

Desjardins, E., & Azevedo, E. (2013). *Food Literacy for Life " Making something out of nothing " Food literacy among youth, young pregnant women*. 1–92.

Diamond, J. J. (2007). Development of a reliable and construct valid measure of nutritional literacy in adults. *Nutrition Journal*, *6*, 4–7. https://doi.org/10.1186/1475-2891-6-5

Doustmohammadian, A., Omidvar, N., & Keshavarz-mohammadi, N. (2018). Developing and validating a scale to measure Food and Nutrition Literacy in elementary school children in Iran. *PLOS One*, 1–18. https://www.researchgate.net/publication/318381551_Developing_and_validating_a_scale_to_measure_Food_and_Nutrition_Literacy_FNLIT_in_elementary_school_children_in_Iran

Doustmohammadian, A., Omidvar, N., Keshavarz-Mohammadi, N., Eini-Zinab, H., Amini, M., & Abdollahi, M. (2022). The association and mediation role of Food and Nutrition Literacy (FNLIT) with eating behaviors, academic achievement and overweight in 10–12 years old students: a structural equation modeling. *Nutrition Journal*, *21*(1), 1–16.

https://doi.org/10.1186/s12937-022-00796-8

Durmus, H., Gökler, M. E., & Havlioglu, S. (2019). Reliability and validity of the Turkish version of the short food literacy questionnaire among university students. *Progress in Nutrition*, *21*(2), 333–338. https://doi.org/10.23751/pn.v21i2.7094

Fingland, D., Thompson, C., & Vidgen, H. A. (2021). Measuring food literacy: Progressing the development of an international food literacy survey using a content validity study. *International Journal of Environmental Research and Public Health*, *18*(3), 1–17. https://doi.org/10.3390/ijerph18031141

Fisher, H., & Potgieter, M. (2024). The development, psychometric properties and refinement of a food literacy scale for specific and general application. *International Journal of Gastronomy and Food Science*, *35*(August 2023), 100862. https://doi.org/10.1016/j.ijgfs.2023.100862

Gibbs, H. D., Camargo, J. M. T. B., Owens, S., Gajewski, B., & Cupertino, A. P. (2018). Measuring Nutrition Literacy in Spanish-Speaking Latinos: An Exploratory Validation Study. *Journal of Immigrant and Minority Health*, *20*(6), 1508–1515. https://doi.org/10.1007/s10903-017-0678-1

Gibbs, H. D., Ellerbeck, E. F., Befort, C., Gajewski, B., Kennett, A. R., Yu, Q., Christifano, D., & Sullivan, D. K. (2016). Measuring Nutrition Literacy in Breast Cancer Patients: Development of a Novel Instrument. *Journal of Cancer Education*, *31*(3), 493–499. https://doi.org/10.1007/s13187-015-0851-y

Gibbs, H. D., Harvey, S., Owens, S., Boyle, D., & Sullivan, D. K. (2017). Engaging experts and patients to refine the nutrition literacy assessment instrument. *BMC Nutrition*, *3*(1). https://doi.org/10.1186/s40795-017-0190-y

Gréa Krause, C., Beer-Borst, S., Sommerhalder, K., Hayoz, S., & Abel, T. (2018). A short food literacy questionnaire (SFLQ) for adults: Findings from a Swiss validation study.

Appetite, *120*, 275–280. https://doi.org/10.1016/j.appet.2017.08.039

Groufh-jacobsen, S., & Medin, A. C. (2023). Food literacy competencies in youth – a mini-review. *Frontiers*, *July*, 1–7. https://doi.org/10.3389/fpubh.2023.1185410

Guiné, R. P. F. F., Florença, S. G., Aparício, G., Cardoso, A. P., & Ferreira, M. (2023). Food Literacy Scale: Validation through Exploratory and Confirmatory Factor Analysis in a Sample of Portuguese University Students. *Nutrients*, *15*(1), 166. https://doi.org/10.3390/nu15010166

Guttersrud, O., Dalane, J. O., & Pettersen, S. (2014). Improving measurement in nutrition literacy research using Rasch modelling: Examining construct validity of stage-specific "critical nutrition literacy" scales. *Public Health Nutrition*, *17*(4), 877–883. https://doi.org/10.1017/S1368980013000530

Hebden, L., Chan, H. N., Louie, J. C., Rangan, A., & Allman-Farinelli, M. (2015). You are what you choose to eat: Factors influencing young adults' food selection behaviour. *Journal of Human Nutrition and Dietetics*, *28*(4), 401–408. https://doi.org/10.1111/jhn.12312

Iskender, H., Yildiz, N., & Dokumacioglu, E. (2024). Investigation of the healthy nutrition attitude and nutritional literacy level of university students receiving health education. *MEDITERRANEAN JOURNAL OF NUTRITION AND METABOLISM*, *17*(1), 35–42. https://doi.org/10.3233/MNM-230050

Jo, E. bin, Yoo, H. lim, Kim, K., & Park, S. (2021). Defining Food Literacy and Its Application to Nutrition Interventions: A scoping Review. *Korean Journal of Community Nutrition*, *26*(2), 77. https://doi.org/10.5720/kjcn.2021.26.2.77

Kabasakal-Cetin, A., Aksaray, B., & Sen, G. (2024). The role of food literacy and sustainable and healthy eating behaviors in ultra-processed foods consumption of undergraduate students. *Food Quality and Preference*, *119*(February), 105232.

https://doi.org/10.1016/j.foodqual.2024.105232

Khorramrouz, F., Doustmohammadian, A., Amini, M., Pourhosein Sarivi, S., Khadem-Rezaiyan, M., Shadmand Foumani Moghadam, M. R., & Khosravi, M. (2022). Validity of a modified food and nutrition literacy questionnaire in primary school children in Iran. *British Journal of Nutrition*, *127*(10), 1588–1597. https://doi.org/10.1017/S0007114521001586

Krause, C., Sommerhalder, K., & Beer-Borst, S. (2016). Nutrition-specific health literacy: development and testing of a multi-dimensional questionnaire. *Ernahrungs Umschau*, *63*(11), 214–220. https://doi.org/10.4455/eu.2016.046

Krause, C., Sommerhalder, K., Beer-Borst, S., & Abel, T. (2016). Just a subtle difference? Findings from a systematic review on definitions of nutrition literacy and food literacy. *Health Promotion International*, *33*(3), 378–389. https://doi.org/10.1093/heapro/daw084

Lai, I.-J. J., Chang, L.-C. C., Lee, C.-K. K., & Liao, L.-L. L. (2021). Nutrition literacy mediates the relationships between multi-level factors and college students' healthy eating behavior: Evidence from a cross-sectional study. *Nutrients*, *13*(10), 3451. https://doi.org/10.3390/nu13103451

Lee, S. Y. D., Stucky, B. D., Lee, J. Y., Rozier, R. G., & Bender, D. E. (2010). Short assessment of health literacy-Spanish and english: A comparable test of health literacy for spanish and english speakers. *Health Services Research*, *45*(4), 1105–1120. https://doi.org/10.1111/j.1475-6773.2010.01119.x

Lee, Y., Kim, T., & Jung, H. (2022). Effects of University Students' Perceived Food Literacy on Ecological Eating Behavior towards Sustainability. *Sustainability (Switzerland)*, *14*(9), 1–16. https://doi.org/10.3390/su14095242

Liao, L.-L. L., Lai, I.-J. J., Shih, S.-F. F., & Chang, L.-C. C. (2018). Development and validation of the nutrition literacy measure for taiwanese college students. *Taiwan*

Journal of Public Health, *37*(5), 582–597. https://doi.org/10.6288/TJPH201810-37(5).107054

Liu, T., Su, X., Li, N., Sun, J., Ma, G., & Zhu, W. (2021). Development and validation of a food and nutrition literacy questionnaire for Chinese school-age children. *PLoS ONE*, *16*(1 January), 1–15. https://doi.org/10.1371/journal.pone.0244197

Liu, Y., Zhang, L., Xu, K., Ding, Y., Li, F., & Zhang, T. (2024). Cross-cultural adaptation and validation of the short nutritional literacy scale for young adults (18-35years) and analysis of the influencing factors. *BMC Public Health*, *24*, 1–12. https://doi.org/https://doi.org/10.1186/s12889-024-19686-1

Luque, B., Villaécija, J., Ramallo, A., de Matos, M. G., Castillo-Mayén, R., Cuadrado, E., & Tabernero, C. (2022). Spanish Validation of the Self-Perceived Food Literacy Scale: A Five-Factor Model Proposition. *Nutrients*, *14*(14), 1–12. https://doi.org/10.3390/nu14142902

Makiabadi, E., Asadollahi, A., Kaveh, M. H., & Salehi, M. (2019). Psychometric Properties of Persian Version of Nutrition Literacy Inventory (NLI-28, 2017) among University Students. *International Journal of Nutrition Sciences*, *4*(3), 113–121. https://doi.org/10.30476/IJNS.2019.82707.1026

McNamara, J., Kunicki, Z. J., Neptune, L., Parsons, K., & Byrd-Bredbenner, C. (2022). Development and Validation of the Young Adult Nutrition Literacy Tool. *Journal of Nutrition Education and Behavior*, *54*(7), 691–701. https://doi.org/10.1016/j.jneb.2022.01.011

McNamara, J., Kunicki, Z. J., Olfert, M. D., Byrd-Bredbenner, C., & Greene, G. (2020). Revision and Psychometric Validation of a Survey Tool to Measure Critical Nutrition Literacy in Young Adults. *Journal of Nutrition Education and Behavior*, *52*(7), 726–731. https://doi.org/10.1016/j.jneb.2020.01.015

Mo, G., Han, S., Gao, T., Sun, Q., Zhang, M., & Liu, H. (2022). Development and validation of a novel short-form nutrition literacy measurement tool for Chinese college students. *Frontiers in Public Health*, *10*(2), 962371. https://doi.org/10.3389/fpubh.2022.962371

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, *27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M., & de Vet, H. C. W. (2014). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties. *Quality of Life Research*, *63*(7), 32.

Morgan, M., Arrowood, J., Farris, A., & Griffin, J. (2023). Assessing food security through cooking and food literacy among students enrolled in a basic food science lab at Appalachian State University. *Journal of American College Health*, *71*(1), 30–35. https://doi.org/10.1080/07448481.2021.1880414

Mostafazadeh, P., Jafari, M. J., Mojebi, M. R., Nemati-Vakilabad, R., & Mirzaei, A. (2024). Assessing the relationship between nutrition literacy and eating behaviors among nursing students: a cross-sectional study. *BMC Public Health*, *24*(1), 1–12. https://doi.org/10.1186/s12889-023-17468-9

Na MS, Y., Cho PhD, M. S., Na, Y., & Cho, M. S. (2020). Development of a tool for food literacy assessment for young adults: Findings from a Korean validation study. *Asia Pacific Journal of Clinical Nutrition*, *29*(4), 876–882. https://doi.org/10.6133/apjcn.202012_29(4).0024

Neuhauser, L., Rothschild, R., & Rodríguez, F. M. (2007). MyPyramid.gov: Assessment of Literacy, Cultural and Linguistic Factors in the USDA Food Pyramid Web Site. *Journal*

of Nutrition Education and Behavior, *39*(4), 219–225. https://doi.org/10.1016/j.jneb.2007.03.005

Nielsen-Bohlman, L., Panzer, A. M., & Kindig, D. A. (2004). *Health Literacy: A Prescription to End Confusion* (L. Nielsen-Bohlman, A. M. Panzer, & D. A. Kindig (eds.)). National Academies Press (US). https://doi.org/10.17226/10883

Nutbeam, D. (2000). Health Literacy. *Health Promotion International*, *15*(3), 259–267. https://doi.org/10.1007/978-3-030-83469-2_19

Page, M. J., McKenzie, J. E., Bossuyt, P., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *Medicina Fluminensis*, *57*(4), 444–465. https://doi.org/10.21860/medflum2021_264903

Palumbo, R., Annarumma, C., Adinolfi, P., Vezzosi, S., Troiano, E., Catinello, G., & Manna, R. (2017). Crafting and applying a tool to assess food literacy: Findings from a pilot study. *Trends in Food Science and Technology*, *67*, 173–182. https://doi.org/10.1016/j.tifs.2017.07.002

Papadaki, A., Hondros, G., A. Scott, J., & Kapsokefalou, M. (2007). Eating habits of University students living at, or away from home in Greece. *Appetite*, *49*(1), 169–176. https://doi.org/https://doi.org/10.1016/j.appet.2007.01.008

Perry, E. A., Thomas, H., Samra, H. R., Edmonstone, S., Davidson, L., Faulkner, A., Petermann, L., Manafò, E., & Kirkpatrick, S. I. (2017). Identifying attributes of food literacy: A scoping review. *Public Health Nutrition*, *20*(13), 2406–2415. https://doi.org/10.1017/S1368980017001276

Poelman, M. P., Dijkstra, S. C., Sponselee, H., Kamphuis, C. B. M., Battjes-Fries, M. C. E.,

Gillebaart, M., & Seidell, J. C. (2018). Towards the measurement of food literacy with respect to healthy eating: The development and validation of the self perceived food literacy scale among an adult sample in the Netherlands. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), 1–12. https://doi.org/10.1186/s12966-018-0687-z

Ramírez-Contreras, C., Farrán-Codina, A., Izquierdo-Pulido, M., & Zerón-Rugerio, M. F. (2021). A higher dietary restraint is associated with higher BMI: a cross-sectional study in college students. *Physiology and Behavior*, *240*(May). https://doi.org/10.1016/j.physbeh.2021.113536

Reynolds, J. S., Treu, J. A., Njike, V., Walker, J., Smith, E., Katz, C. S., & Katz, D. L. (2012). The Validation of a Food Label Literacy Questionnaire for Elementary School Children. *Journal of Nutrition Education and Behavior*, *44*(3), 262–266. https://doi.org/10.1016/j.jneb.2011.09.006

Rhea, K. C., Cater, M. W., McCarter, K., & Tuuri, G. (2020). Psychometric Analyses of the Eating and Food Literacy Behaviors Questionnaire with University Students. *Journal of Nutrition Education and Behavior*, *52*(11), 1008–1017. https://doi.org/10.1016/j.jneb.2020.05.002

Ringland, E. M., Gifford, J. A., Denyer, G. S., Thai, D., Franklin, J. L., Stevenson, M. M., Prvan, T., & O'connor, H. T. (2016). Evaluation of an electronic tool to assess food label literacy in adult Australians: A pilot study. *Nutrition and Dietetics*, *73*(5), 482–489. https://doi.org/10.1111/1747-0080.12271

Rosas, R., Pimenta, F., Leal, I., & Schwarzer, R. (2020). FOODLIT-PRO: Food literacy domains, influential factors and determinants—A qualitative study. *Nutrients*, *12*(1), 1–31. https://doi.org/10.3390/nu12010088

Rothman, R. L., Housam, R., Weiss, H., Davis, D., Gregory, R., Gebretsadik, T., Shintani,

A., & Elasy, T. A. (2006). Patient Understanding of Food Labels. The Role of Literacy and Numeracy. *American Journal of Preventive Medicine*, *31*(5), 391–398. https://doi.org/10.1016/j.amepre.2006.07.025

Schaafsma, H. N., Caruso, O. T., Mceachern, L. W., & Seabrook, J. A. (2025). ARTICLE IN PRESS Research Article Understanding Food Literacy Intervention Effectiveness : Postsecondary Students ' Perspectives on How a mHealth Food Literacy Intervention Impacted Their Dietary Behaviors. *Journal of Nutrition Education and Behavior*, *000*(000). https://doi.org/10.1016/j.jneb.2025.01.003

Silk, K. J., Sherry, J., Winn, B., Keesecker, N., Horodynski, M. A., & Sayir, A. (2008). Increasing Nutrition Literacy: Testing the Effectiveness of Print, Web site, and Game Modalities. *Journal of Nutrition Education and Behavior*, *40*(1), 3–10. https://doi.org/10.1016/j.jneb.2007.08.012

Silva, P., Araújo, R., Lopes, F., & Ray, S. (2023). Nutrition and Food Literacy: Framing the Challenges to Health Communication. *Nutrients*, *15*(22), 1–25. https://doi.org/10.3390/nu15224708

Slater, J. (2013). Is cooking dead? The state of Home Economics Food and Nutrition education in a Canadian province. *International Journal of Consumer Studies*, *37*(6), 617–624. https://doi.org/10.1111/ijcs.12042

Slater, J., Falkenberg, T., Rutherford, J., & Colatruglio, S. (2018). Food literacy competencies: A conceptual framework for youth transitioning to adulthood. *International Journal of Consumer Studies*, *42*(5), 547–556. https://doi.org/10.1111/ijcs.12471

Smith, M. gale. (2009). Unexamined Food Is Not Worth Eating : Food Literacy As the Foundation for Food and Nutrition Courses. *Conference Paper*, *March 2009*.

Sprake, E. F., Russell, J. M., Cecil, J. E., Cooper, R. J., Grabowski, P., Pourshahidi, L. K., &

Barker, M. E. (2018). Dietary patterns of university students in the UK: A cross-sectional study. *Nutrition Journal*, *17*(1), 1–17. https://doi.org/10.1186/s12937-018-0398-y

Stjernqvist, N. W., Elsborg, P., Ljungmann, C. K., Benn, J., & Bonde, A. H. (2021). Development and validation of a food literacy instrument for school children in a Danish context. *Appetite*, *156*(March 2020), 1–10. https://doi.org/10.1016/j.appet.2020.104848

Tabacchi, G., Battaglia, G., Alesi, M., Paoli, A., Palma, A., & Bellafiore, M. (2020). Food literacy predictors and associations with physical and emergent literacy in pre-schoolers: Results from the Training-to-Health Project. *Public Health Nutrition*, *23*(2), 356–365. https://doi.org/10.1017/S1368980019002404

Thomas, H., Azevedo Perry, E., Slack, J., Samra, H. R., Manowiec, E., Petermann, L., Manafò, E., & Kirkpatrick, S. I. (2019). Complexities in Conceptualizing and Measuring Food Literacy. *Journal of the Academy of Nutrition and Dietetics*, *119*(4), 563–573. https://doi.org/10.1016/j.jand.2018.10.015

Thompson, C., Adams, J., & Vidgen, H. A. (2021). Are we closer to international consensus on the term 'food literacy'?: A systematic scoping review of its use in the academic literature (1998–2019). *Nutrients*, *13*(6). https://doi.org/10.3390/nu13062006

Thompson, C., Byrne, R., Adams, J., & Vidgen, H. A. (2022). Development, validation and item reduction of a food literacy questionnaire (IFLQ-19) with Australian adults. *International Journal of Behavioral Nutrition and Physical Activity*, *19*(1), 1–23. https://doi.org/10.1186/s12966-022-01351-8

Truman, E., Bischoff, M., & Elliott, C. (2019). Which literacy for health promotion : health , food , nutrition or media ? *Health Promotion International*, *35*(February 2019), 432–444. https://doi.org/10.1093/heapro/daz007

Truman, E., Lane, D., & Elliott, C. (2017). Defining food literacy: A scoping review.

Appetite, *116*(2017), 365–371. https://doi.org/10.1016/j.appet.2017.05.007

Velardo, S. (2015). The Nuances of Health Literacy, Nutrition Literacy, and Food Literacy. *Journal of Nutrition Education and Behavior*, *47*(4), 385-389.e1. https://doi.org/10.1016/j.jneb.2015.04.328

Vettori, V., Lorini, C., Milani, C., & Bonaccorsi, G. (2019). Towards the Implementation of a Conceptual Framework of Food and Nutrition Literacy: Providing Healthy Eating for the Population. *International Journal of Environmental Research and Public Health*, *16*(24). https://doi.org/https://doi.org/10.3390/ijerph16245041

Vidgen, H. A., & Gallegos, D. (2014). Defining food literacy and its components. *Appetite*, *76*(2014), 50–59. https://doi.org/10.1016/j.appet.2014.01.010

Vrinten, J., Royen, K. Van, Pabian, S., Backer, C. De, & Matthys, C. (2023). Development and validation of a short nutrition literacy scale for young adults. *Frontiers in Nutrition*, *March*, 1–10. https://doi.org/10.3389/fnut.2023.1008971

Wallace, R., Lo, J., & Devine, A. (2016). Tailored nutrition education in the elderly can lead to sustained dietary behaviour change. *Journal of Nutrition, Health and Aging*, *20*(1), 8–15. https://doi.org/10.1007/s12603-016-0669-2

Watson, W. L., Chapman, K., King, L., Kelly, B., Hughes, C., Louie, J. C. Y., Crawford, J., & Gill, T. P. (2013). How well do Australian shoppers understand energy terms on food labels? *Public Health Nutrition*, *16*(3), 409–417. https://doi.org/10.1017/S1368980012000900

Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., Mockbee, J., & Hale, F. A. (2005). Quick assessment of literacy in primary care: The newest vital sign. *Annals of Family Medicine*, *3*(6), 514–522. https://doi.org/10.1370/afm.405

Whatnall, M. C., Patterson, A. J., Burrows, T. L., & Hutchesson, M. J. (2019). Higher diet

quality in university students is associated with higher academic achievement: a cross-sectional study. *Journal of Human Nutrition and Dietetics*, *32*(3), 321–328. https://doi.org/10.1111/jhn.12632

Wijayaratne, S., Westberg, K., Reid, M., & Worsley, A. (2022). Developing food literacy in young children in the home environment. *International Journal of Consumer Studies*, *46*(4), 1165–1177. https://doi.org/10.1111/ijcs.12750

Yan, W., Caihong, H., Xuefeng, Y., & Jiayu, Z. (2023). Evaluation of the nutrition literacy assessment questionnaire for college students and identification of the influencing factors of their nutrition literacy. *BMC Public Health*, *23*(1), 1–11. https://doi.org/10.1186/s12889-023-17062-z

Yuen, E. Y. N., Thomson, M., & Gardiner, H. (2018). Measuring Nutrition and Food Literacy in Adults: A Systematic Review and Appraisal of Existing Measurement Tools. *Health Literacy Research and Practice*, *2*(3), 134–160. https://doi.org/https://doi.org/10.3928/24748307-20180625-01

Zeminian, L. B., Corona, L. P., Batista, I. do N., da Silva, M. C., & da Cunha, D. T. (2022). Translation, Adaptation, and Validity of the Short Food Literacy Questionnaire for Brazil. *Foods*, *11*(24), 1–12. https://doi.org/10.3390/foods11243968

Zhang, Y., Sun, Q., Zhang, M., Mo, G., & Liu, H. (2022). *Nutrition Literacy Measurement Tool With Multiple Features for Chinese Adults*. *43*(2), 189–200. https://doi.org/10.1177/03795721211073221

Zwierczyk, U., Kobryn, M., & Duplaga, M. (2022). Validation of the Short Food Literacy Questionnaire in the Representative Sample of Polish Internet Users. *International Journal of Environmental Research and Public Health*, *19*(15). https://doi.org/10.3390/ijerph19159710

Appendices

Appendix A: COSMIN risk of bias checklist

Tool Author (Year)	Concept												
	PROM* development	Structural/Construct Validity	Internal Consistency	Cross-cultural Validity	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing		Responsiveness			
								Convergent Validity	Discriminative or Known Groups Validity	Gold Standard	Other Outcome	Between Subgroup	Intervention
NL-TCS Liao (2018)	A	V	V	NR	NR	NR	NR	V	NR	NR	NR	NR	NR
Turkish SFLQ Durmus (2019)	NA	A	V	NR	D	NR	NR	V	V	NR	NR	NR	NR
NLI-28 Makiabadi (2019)	NA	V	V	NR	NR	NR	I	I	D	NR	NR	NR	NR
CNLT-R McNamara (2020)	NA	V	D	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
KFLS Na & Cho (2020)	V	V	V	NR	V	NR	NR	NR	NR	NR	NR	NR	NR
EFLBQ Rhea et al (2020)	V	V	V	NR	D	NR	NR	NR	NR	NR	NR	NR	NR
NL-SF12 Mo (2022)	NA	V	V	NR	NR	NR	V	NR	NR	NR	NR	NR	NR
Spanish SPFL Luque (2022)	NA	V	V	NR	NR	NR	NR	V	V	NR	NR	NR	NR
Korean SPFL Lee, Y. (2022)	NA	V	V	NR	NR	NR	NR	V	V	NR	NR	NR	NR
YA-NLT McNamara (2022)	V	V	V	NR	NR	NR	NR	V	NR	NR	NR	NR	NR
FNLI Demir & Özer (2022)	V	V	V	NR	V	NR	NR	NR	V	NR	NR	NR	NR
FLS Guine (2023)	D	V	V	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

NLAQ Yan (2023)	NA	V	V	NR	NR	NR	NR	NR	V	NR	NR	NR	NR
S-NutLit Vrinten (2023)	V	A	V	NR	V	NR	NR	V	V	NR	NR	NR	NR
Iranian NL-SF12 Mostafazadeh (2024)	NA	V	V	NR	V	NR	NR	V	V	NR	NR	NR	NR
CS-NutLit Liu (2024)	NA	V	V	NR	V	NR	NR	NR	V	NR	NR	NR	NR

*PROM (Patient-related outcome measurement) development box according to COSMIN checklist was applied to newly developed tools; Boxes 1 (PROM development) and 2 (Content validity) were integrated into one box and modified due to inapplicability of clinical setting outcome measures to our study. The scores for this domain was not reported in the systematic review, but content validity was reported narratively in corresponding sections.

NL-TCS, Nutrition Literacy for Taiwanese College students; Turkish SFLQ, Turkish Short Food Literacy Questionnaire; NLI-28, Nutrition Literacy Inventory – 28 items; CNLT-R, Critical Nutrition Literacy Tool – Revised; KFLS, Korean Food Literacy Scale; EFLBQ, Eating and Food Literacy Behaviors Questionnaire; NL-SF12, Nutrition Literacy – Short Form 12 items; Spanish SPFL, Spanish Self-perceived Food Literacy; Korean SPFL, Korean Self-perceived Food Literacy; YA-NLT, Young Adults – Nutrition Literacy Tool; FNLI, Food and Nutrition Literacy Instrument; FLS, Food Literacy Scale; NLAQ, Nutrition Literacy Assessment Questionnaire; S-NutLit, Short-Nutrition Literacy; Iranian NL-SF12, Iranian Nutrition Literacy – Short Form 12 items; CS-NutLit, Chinese Short – Nutrition Literacy.

V, Very Good; A, Adequate; D, Doubtful; I, Inadequate; NR, Not Reported; NA, Not Applicable. The ‘structural validity’ changed to ‘structural/construct validity’ due to being a subset of construct validity.

Appendix B: Utility frequency of measurement tools

Tool name	Approximate number of usages	Frequency, %
SFLQ	10: (Zwierczyk, 2022), (Zeminian, 2022), (Duan, 2022), (Itzkovitz, 2022), (Hoteit, 2023), (Bookari, 2023), (Zeminian, 2024), (Saengrut, 2024), (Sercu, 2024), (Uzdil, 2024)	27.78
SPFL	9: (Sponselee, 2021), (Boslooper-Meulenbelt, 2021), (Polhuis, 2023), (Murakami, 2024), (Greenberg, 2025), (Çelik, 2024), (Griebler, 2024), (Jacobs, 2024), (Murakami, 2025)	25.0
NL-TCS	5: (Lee CK, 2019), (Liao, 2019), (Lai IJ, 2021), (Banna, 2022), (Lai IJ, 2023)	13.89
EFLBQ	4: (Rhea, 2021), (Ludwig, 2024), (Murphrey, 2024), (Carr, 2024)	11.11
S-NutLit	2: (Chilón-Troncos, 2024), (Koc, 2024)	5.56
FLS	2: (Guine, 2023), (Boariu, 2024)	5.56
NL-SF12	1: (Mo, 2024)	2.78
NLAQ	1: (Liu, 2021)	2.78
NLI-28	1: (Nikmah, 2023)	2.78
CNLT-R	1: (Bedoyan, 2020)	2.78
YA-NLT	0	0
FNLI	0	0
KFLS	0	0

Note* The original scales on which the adapted ones were based were used to measure the frequency of tool utility across the field (e.g., SFLQ, SPFL, NL-SF12, S-NutLit).

Appendix C: Search strategy for the systematic review

Database	Search Strategy
PubMed	(("food literacy"[tiab] OR "nutrition literacy"[tiab] OR "nutritional literacy"[tiab]) AND ("college student*"[tiab] OR "university student*"[tiab] OR "undergraduat*"[tiab] OR "postsecondary student*"[tiab]) AND ("measure*"[tiab] OR "questionnaire"[tiab] OR "scale"[tiab] OR "tool"[tiab] OR "instrument"[tiab] OR "validity"[tiab] OR "reliability"[tiab] OR "framework"[tiab] OR "conceptual*"[tiab]))
MEDLINE	(TI "food literacy" OR TI "nutrition literacy" OR TI "nutritional literacy" OR AB "food literacy" OR AB "nutrition literacy" OR AB "nutritional literacy") AND (TI "college student*" OR TI "university student*" OR TI "undergraduat*" OR AB "college student*" OR AB "university student*" OR AB "undergraduat*") AND (AB "measure*" OR AB "questionnaire*" OR AB "scale*" OR AB "instrument*" OR AB "validity" OR AB "reliability" OR AB "framework" OR AB "conceptualization")
ProQuest	("food literacy" OR "nutrition literacy" OR "nutritional literacy") AND ("college student*" OR "university student*" OR "undergraduat*" OR "postsecondary student*") AND ("measurement tool*" OR "questionnaire*" OR "scale*" OR "instrument*" OR "validity" OR "reliability" OR "framework" OR "conceptualization")
Web of Science	TS=("food literacy" OR "nutrition literacy" OR "nutritional literacy") AND TS=("college student*" OR "university student*" OR "undergraduat*" OR "postsecondary student*") AND TS=("measure*" OR "questionnaire*" OR "scale*" OR "tool*" OR "instrument*" OR "validity" OR "reliability" OR "framework" OR "conceptual*")
CINAHL	(TI "food literacy" OR TI "nutrition literacy" OR TI "nutritional literacy" OR AB "food literacy" OR AB "nutrition literacy" OR AB "nutritional literacy") AND (TI "college student*" OR TI "university student*" OR TI "undergraduat*" OR AB "college student*" OR AB "university student*" OR AB "undergraduat*") AND (AB "measure*" OR AB "questionnaire*" OR AB "scale*" OR AB "instrument*" OR AB "validity" OR AB "reliability" OR AB "framework" OR AB "conceptualization")