



SUSTAINABILITY MEASUREMENT SCALES: CRITERIA, SPECIFICITIES, AND BASIC PROTOCOL

ESCALAS DE MENSURAÇÃO DA SUSTENTABILIDADE: CRITÉRIOS, ESPECIFICIDADES E PROTOCOLO BÁSICO

EDSON LUIS KUZMA

Doutor em Administração pela Universidade do Oeste de Santa Catarina - UNOESC (2022). Doutorando bolsista pela Capes/Prosup. Mestre em Desenvolvimento Comunitário pela Universidade Estadual do Centro-Oeste - UNICENTRO (2016). Especialista em Gestão Pública pela Universidade Estadual do Centro-Oeste - UNICENTRO (2018). Bacharel em Administração pela Universidade Estadual do Centro-Oeste - UNICENTRO (2014). Bacharel em Ciências Contábeis pela Universidade Paranaense - UNIPAR (2022). ORCID Id: <http://orcid.org/0000-0001-7784-1287>.

SIMONE SEHNEM

Pós- Doutorado pela FGV com período de coleta de dados na Escócia e na Inglaterra (2016 à 2018). Doutorado em Administração e Turismo pela Univali/SC (2011). Mestrado em Administração pela Universidade Federal de Santa Catarina (2007). Graduação em Agronegócios pela Universidade do Oeste de Santa Catarina - Campus de São Miguel do Oeste (2005) e graduação em Administração pela Universidade do Oeste de Santa Catarina (2010). É Bolsista Produtividade PQ-2 Cnpq. Atua como Professora e Pesquisadora na UNOESC e na UNISUL. Na Universidade do Oeste de Santa Catarina (UNOESC) no Doutorado Acadêmico em Administração e no Mestrado Profissional em Administração nos quais desenvolve estudos nas Linhas de Pesquisa Sustentabilidade, empreendedorismo e dinâmicas territoriais e Sustentabilidade e Organizações. Ministra as disciplinas na área de sustentabilidade.

MARCELO LUIZ DIAS DA SILVA GABRIEL

Doutor em Educação (Ciência e Tecnologia) pela Universidade Estadual de Campinas - UNICAMP (2011), Mestre em Administração de Empresas pelo Centro Universitário Álvares Penteado - UNIFECAP (2005), MBA Executivo em Marketing pelo IBMEC (2001) e Bacharel em História pela Universidade de São Paulo - USP (1997). Professor do Programa de Pós-Graduação em Administração e do Programa de Mestrado Profissional em Comportamento do Consumidor da Escola Superior de Propaganda e Marketing (ESPM). Atua como revisor de periódicos nacionais e internacionais nas áreas de Administração e Psicologia, e como parecerista de eventos como EnANPAD, CLAV, SEMEAD e ENGEMA. Editor convidado da Brazilian Journal of Marketing (BJMkt) na edição especial sobre Pesquisa de Marketing. Pesquisa e publica nos seguintes temas: Comportamento do Consumidor B2B, Gestão do Marketing Verde, Sustentabilidade e Inovação, Produção e Consumo Sustentável, Métodos Quantitativos e Metodologia de Pesquisa. Membro do Comitê Editorial - Brazilian Journal of Marketing (BJMkt). Atuou como executivo em multinacionais nas áreas de Vendas e Marketing.





ERIC DAVID COHEN

Graduado em Engenharia Elétrica pela Universidade S. Judas Tadeu (1982), com Mestrado em Administração pela Clemson University (1997) e Master of Computer Science pela Arizona State University (1993). Doutor em Administração de Empresas pela Fundação Getulio Vargas (2004) e Pós-doutorado em Administração pela Universidade Mackenzie.

MACIEL MANOEL DE QUEIRÓZ

Possui Doutorado (2017) e Mestrado (2011) em Engenharia Naval e Oceânica pela Escola Politécnica da Universidade de São Paulo, Especialização em Logística pela Universidade Presbiteriana Mackenzie (2008) e Graduação em Administração de Empresas pelo Centro Universitário de Santo André (2006).

ABSTRACT

Sustainability is an essential construct for social and economic development. It requires an objective definition of measurement methods and helps decision-makers clarify and discern their choices. This article aims to map and analyze the specifics of measurement scales with a scope aimed at sustainability to advance the study of sustainability measures. The final sample consists of 31 articles collected to i) identify analytical elements regarding the specificities of the scales; ii) check their scope; iii) evaluate the reliability and validity measures used; iv) map the main limitations; and v) propose a sustainability research agenda. The study contributes by categorizing and analyzing the measurement process. Furthermore, it makes it possible to understand the stages necessary for elaborating and validating measurement instruments specific to this field of research.

Keywords: sustainability, measurement scales, sustainability metrics, development of scales.

RESUMO

A sustentabilidade é um construto importante para o desenvolvimento social e econômico. Requer uma definição objetiva dos métodos de medição e ajuda os tomadores de decisão a esclarecer e discernir suas escolhas. Este artigo tem como objetivo mapear e analisar as especificidades de escalas de mensuração com escopo voltado para a sustentabilidade, com o objetivo de avançar no estudo de medidas de sustentabilidade. A amostra final foi composta por 31 artigos coletados com o objetivo de: i) identificar elementos analíticos referentes às especificidades das escalas; ii) verificar seu escopo; iii) avaliar as medidas de confiabilidade e validade utilizadas; iv) mapear as principais limitações; e v) propor uma agenda de pesquisa em sustentabilidade. O estudo contribui categorizando e analisando o processo de medição. Possibilita compreender as etapas necessárias para a elaboração e validação de instrumentos de medida específicos para esse campo de pesquisa.

Palavras-chave: Sustentabilidade, escalas de mensuração, métricas de sustentabilidade, desenvolvimento de escalas.





1 INTRODUCTION

Although their roots can be found in studies from the 19th century (Lumley & Armstrong, 2004), theories of sustainability are an emerging field. Consequently, they remain to be consolidated as an area of knowledge. The central definitions and measures that involve the theme include extensions in different directions and approaches to research. Recent advances in sustainability are associated, among others, with the field of sustainable development (Asokan, Yarime, and Onuki, 2019), a topic that revisits discussions fostered over many decades about the limits of growth and the planet's ability to absorb anthropic action (Carson, 1962; Meadows, Meadows, Randers and Behrens, 1972). Given this context, it is relevant for companies and researchers to identify and develop practices and strategies to meet current needs without negatively compromising future possibilities of existence (WCED, 1987).

Advances in practices, strategies, behavior change, and the establishment of sustainability management policies must be mapped and understood by forming indicators, metrics, and analysis parameters. Strengthening the field of sustainability involves the arduous and complex task of constructing theoretical and empirical considerations that allow the design and validation of measurement models. The manifestation in different aspects of practices associated with sustainability can be better understood when measured by indicators. The interpretation of quantifiable elements allows the field to advance on promising fronts for organizations and society, mainly due to achieving mutual gains through improving practices.

The justification for carrying out the study is associated with the need to expand knowledge about the field of sustainability and generate possibilities for further advances. Environmental issues and sustainability progress are essential for many disciplines (Nilashi et al., 2019). However, to date, studies have yet to be identified that sought to develop a systematic literature review on valid measurement scales applied to sustainability factors in a holistic sense. Some initiatives are moving toward measuring sustainability indicators (Corona et al., 2019), but the issue of building measurement





scales still needs to be clarified. Therefore, this review contributes to scientific development by advancing and compiling forms of measurement instruments identified in different studies that lead to the development of scales. Furthermore, an analysis is provided on which stages are performed and which techniques are used to provide reliability and validity to the instruments in this field. The synthesis of this information can add knowledge to the field of sustainability and guide the execution of future research.

In the context of sustainability metrics, concerns about the validity and reliability of the measures are inserted, which support the necessary and promoted advances in obtaining processes that are more aligned with sustainable practices. Furthermore, exploring the scales allows us to understand the current panorama of the actions effectively implemented and to envision what impacts they can generate in reducing the consumption of virgin resources and other activities associated with social and environmental aspects. As a complex theme explored on different fronts, the scaling proposal refines the assessment of the fundamental advances being made, which is why a systematic review in the field of sustainability is relevant.

Therefore, this article aims to analyze the specificities of measurement scales with a scope aimed at sustainability. From a literature review, analytical categories were proposed to assess the scales in terms of the following aspects: i) type of scale, ii) scale items, iii) type of theoretical construct, iv) analysis technique, v) guidance and research protocol, vi) scale, vii) reliability measures, and viii) validity measures. The articles were reviewed regarding the theoretical domain for which the scale was developed, as well as identifying the stages common to the protocols observed in the development and refinement of the scales. The analytical dimensions were defined based on aspects related to the construction and validation of scales from derivations of scale development paradigms as prescribed by Churchill (1991), DeVellis (2003), Hinkin (1998), Netemeyer, Bearden, and Sharma (2003) and Hair et al. (2009). The study provides an analysis of the main characteristics of scales developed for the field of sustainability. From the reviewed articles and the limitations identified for these items, proposals for future studies related to validating scales in sustainability are outlined.





2 MEASURES AND MEASUREMENTS IN THE FIELD OF SUSTAINABILITY

The concept of sustainability can be described through different structures. In the context of this research, sustainability is understood as sustaining, supporting, conserving, and making natural resources perennial (environmental quality) to permit a dignified life for people (social justice) and economic prosperity capable of promoting the viability of human enterprises (Elkington, 1997). Its measurement depends on criteria and the approach with which the conceptual structure and domain are conceived (Alvarado-Herrera et al., 2017). Before any possibility of use, measurement instruments must provide accurate, valid, and interpretable data to assess sustainability-related factors. In this sense, it is considered essential that the instrument presents sufficiently robust measures so that the results are appropriate (Bonett & Wright, 2014).

A measurement scale is a set of rules that allows numbers or labels to be assigned to objects, people, states, or events to quantify and qualify their attributes. In large part, performance in obtaining results is attributed to the reliability and validity of the instruments. However, validity and reliability depend upon defining the elements that make up the construct, namely, its conceptual domain. Delimiting the theoretical domains implies pointing out a substantial part of the articulation and theoretical construction of the phenomenon. Assigning limits also involves deciding the conceptual breadth of the construction and specifying its dimensions. In this sense, the quality of the responses on a scale, regardless of the field for which it is intended, can be assessed from the use for which it is intended and the fidelity of the measured instrument; this concept is related to the validity and reliability of the instrument (Haynes, Richard and Kubany, 1995).

Creating a new instrument does not guarantee that it has been appropriately validated (Turner & Zolin, 2012). A good piece of advice is that researchers must evaluate the properties of an instrument before it is used (Straub, 1989; Straub & Gefen, 2004). The absence of a protocol prescribed in the literature or even a description of the steps taken for its development renders the veracity of any result precarious. When theoretical inaccuracy compromises measurement, it also confuses the advancement of theoretical domains. In sustainability, where theoretical consolidation is still being sought (Brink, Hengeveld, and Tobi, 2020), consistency of argument and method are considered





essential. Along this line, conducting empirical studies helps researchers build and provide content validity, especially to review errors, and identify poorly conceived theoretical arguments.

The conception of proposals for measuring sustainability demands a stated definition. The most advanced initiatives in this regard are sustainability indicators and metrics. A significant milestone in the definition of sustainability – which aligns concepts based on economic, social, and environmental development theories – is given in the Brundtland Commission report (WCED, 1987). The report prescribes sustainable development to meet present needs without compromising its provision for future generations. Sustainability is also taken from its main dimensions, including the three pillars of environmental, social, and economic concerns (Elkington, 1997).

The measurement of sustainability as an interdisciplinary field of research was developed by Brink, Hengeveld, and Tobi (2020). The authors draw attention to the fact that sustainability measures use different metrics in different disciplines. This dispersion compromises the articulation of different studies when they quantify the field based on conflicting attributes. Although the consensus is generally mild to the approach via the triple bottom line (TBL), the notion of sustainability is used to measure several different objects. Through a systematic review, the authors identified that not all studies that measure sustainability adopt the TBL approach. In the studies, it was also impossible to identify sufficient information about the procedures adopted in the measurements. The inaccuracy of measuring instruments was recorded in the three dimensions of sustainability.

The concept of sustainability measurement is associated with the definition with which the theme is taken, which aligns with the practices adopted to improve and provide more information on the social and environmental impacts caused by the impacts of organizations. Predominantly, the literature on sustainability indicators focuses on environmental and social indicators and how companies incorporate these principles into their management practices (Roca & Searcy, 2012). Sustainability and its mediation are taken from a multidimensional perspective, which in the context of companies involves





articulating management actions at all levels, from operational to strategic (Adams et al., 2016).





3 METHOD

The research execution protocol adopts the elaboration stages proposed by Tranfield, Denyer, and Smart (2003). Two main stages were defined in the selection of articles. In the first stage of the systematic review, all possible studies that address the development and validation of sustainability scales indexed in relevant databases were included. The initial definition of the keywords was obtained through preliminary investigation, followed by the test of word sets related to the theoretical domains of sustainability and validation of scales in the Web of Science, ScienceDirect, and Scopus databases, as described by Bonisoli, Galdeano- Gómez, and Piedra-Muñoz (2018). Based on the preliminary search with the two broad terms, the themes were expanded by incorporating new keywords. The choice of keywords was made to limit the scope of the study to only research in line with the theme, as already performed by Sehnem et al. (2019) and Comin et al. (2019) in their systematic literature reviews. Related terms such as sustainable supply chain, clean production, and clean energy have not been included.

The final bases for consultation were Scopus, Web of Science, ScienceDirect, Sage, Emerald, ProQuest, and Wiley Online Library. Other databases were consulted; however, the articles found were already indexed in the initial databases, so they were not considered. The search was carried out from March 03-05, 2021. No chronological restrictions were used in the selection of studies.

Only journals with peer review were considered to guarantee the quality of the revised material. Inclusion was restricted to articles in the English language. The selected articles were presented in their full versions. The scope was restricted to Business, Management, and Accounting, and a strict relationship was maintained with the review focus. Such delimitation excluded systematic reviews in the set from being analyzed. In the second stage of the systematic review, exclusion criteria for articles were applied to compose the final sample to be analyzed. The general procedures for selecting articles are based on the recommendations of the PRISMA methodology (Moher et al., 2015). The list of studies that comprised the analyzed sample is arranged in the results section. The studies and scales presented were analyzed in detail in the sections.





The exclusion process resulted in selecting only 31 articles as a final sample. Given the narrow scope that conditions the review, the articles suitable for inclusion only use primary data, mainly due to the nature of data extraction and coding. The predominance of articles directs the content of analysis to the broad field of study of corporate sustainability, discussion of internal and external sustainability reports, and analysis on the disclosure of sustainability practices aligned with accounting. Although these studies make numerous and consistent contributions to the field of sustainability study and the promotion of management practices and decisions to managers, they are distant from the review's objective. The approaches mentioned above are abundant in studies that adopt different perspectives but do not contain the elements necessary for the discussion proposed in this article, which refers to the measurement scales. This type of study requires following a specific development protocol, instrument refinement, and validation that differs from other types of study. Therefore, the relevance of the different study approaches is recognized. However, the scope of this article is limited only to those who adopt the process of developing scales, which justifies the inclusion of a reduced number of articles as a final sample for analysis.

4 LITERATURE RESULTS AND STRUCTURAL ASPECTS OF THE REVIEWED ARTICLES

The structure of the reviewed articles is segmented according to the different possibilities of analysis, contemplating the following elements: i) region/continent of the research; ii) type of economy; iii) economic sector; and iv) research approach.

The reviewed studies are concentrated in some regions. Geographic distribution is most concentrated in Asia, with 42% of studies, followed by 32% in Europe, while other regions are less frequent. The analysis refers to the locus of application of the research. We note that an article was developed as a comparative study between New Zealand and the USA (Kapitan, Kennedy, and Berth, 2018). Regarding the type of economy, studies that applied to developed economies, distributed among different continents, such as Europe, America (Robertson & Barling, 2017), and Oceania (Larimian & Sadeghi, 2019), predominate. Of the emerging and underdeveloped economies, all countries are located





in Asia, distributed between Continental Asia (Ahmad, Omar, and Quoquab, 2019), the Middle East (Meiboudi et al., 2017), and Insular Asia (Wang, 2015).

The analysis of the economic sector level shows that the construction and validation of scales in sustainability are developed more frequently in manufacturing (Mani et al., 2016), with 39% of the studies. The application in universities (Biasutti & Frate, 2016) or studies combining manufacturing companies or services (Haan et al., 2018) is also representative. The level of consumers (Taecharungroj, Suksaroj, and Rattanapan, 2018) deals with assessing the perception of these audiences regarding aspects associated with sustainability and sustainable development in different contexts. Finally, an isolated study addresses social organizations (Kannampuzha & Hockerts, 2019).

The analysis of the economic sector level shows that the construction and validation of scales in sustainability are developed more frequently in the context of universities (Biasutti & Frate, 2016), with 32% of the studies. The application in manufacturing (Mani et al., 2016) or in studies that combine manufacturing companies or services (Haan et al., 2018) is also representative. The level of consumers (Taecharungroj, Suksaroj, and Rattanapan, 2018) deals with assessing the perception of these audiences regarding aspects associated with sustainability and sustainable development in different contexts. Finally, an isolated study addresses social organizations (Kannampuzha & Hockerts, 2019).

The research approach aspect indicates a similar proportion between studies that build and validate scales using purely quantitative techniques and those that articulate the quantitative to the qualitative method. Therefore, this process is adopted as a scale refinement and validation technique (Cabrera-Nguyen, 2010).

5 ANALYTICAL DIMENSIONS, SCOPE, RELIABILITY, AND VALIDITY OF SCALES

For the necessary categorization and analysis of the studies, it was decided to use a definition anchored in the dimensions segmented according to elements of an analytical nature (type of scale, scale items, type of construct, analysis technique, orientation), scale scope, the measure of reliability, and measure of validity.





The discussion of the results was categorized in light of the different dimensions established in this classification. All elements identified in the literature consider the following order: 6.1 Type of scale, 6.2 Items of scale (Likert type), 6.3 Type of construct, 6.4 Analysis technique, 6.5 Orientation, 6.6 Range of the scale, 6.7 Measure of reliability, and 6.8 Measure of validity.

5.1 SCALE TYPE

The analyzed dimension of the scale type refers to levels that measure the relationship between numerical values that the researchers attribute during measurement. The scale type identifies how the respondents' information is transformed into data and how it is encoded. The classification used as a basis is that of Stevens (1946), which differentiates four measures: nominal, ordinal, interval, and ratio. Likert variables assume discrete values. Each scale modality generates different data types, which descriptive or inferential statistics can manipulate. The statistical analysis of these data must correspond to the appropriate form of manipulation. Of the reviewed articles, 27 use ordinal-type scales. In the other four studies, it was impossible to identify the scale type used, although the multivariate tests described indicate the ordinal type.

5.2 SCALE ITEMS (LIKERT TYPE)

The reviewed articles present the same specific type of response format. Except for one article (for which it is impossible to identify the format), the others use a Likert scale. The Likert scale (or scale of agreement) is widely adopted in measurement instruments that measure opinions, preferences, and attitudes (DeVellis, 2003). In addition, scales with this format were previously used in studies on sustainability and the environment (Alvarado-Herrera et al., 2017).

The number of assessment items that reflect the assessment of a sentence ranges from four to seven points. The predominance is scaled with a single central or neutral item with an odd number of answer options. There are seven scales with the most significant number of response possibilities (Gupta & Agrawal, 2017), with 13 studies. The 5-point





scales represent the majority of studies (15 articles), being the most used. This modality has a neutral point and two other answer options, which vary in agreement or disagreement. Another three studies adopt scales without neutral or central points, which causes the respondent to be forced to position himself positively or negatively. These articles adopt four (Waltner, Riess, and Mischo, 2019) or six (Khan & Quaddus, 2015) points for evaluating the sentences in the questionnaire. With an even number of options, this type of scale cannot be considered a Likert-type scale as proposed initially (Likert, 1932). The choice of a neutral response is common in surveys conducted in Asian countries (Trompenaars & Hampden-Turner, 1998).

5.3 TYPE OF CONSTRUCT

In specifying a construct's measurement types, the researcher must consider two broad approaches: formative and reflective models (Hair et al., 2009). The most commonly used construction is to define the design of the measurement of the construct from a reflective model (Coltman et al., 2008). In the reflexive construct, the latent variable manifests itself through the items (indicators or manifest variables). The set of manifest variables is coded in the same conceptual direction, or the conceptual construction causes the indicators. In the case of the formative construct, the latent variable is formed from the manifest variables. Individually, the indicators influence and define the shape of the latent construct. Of the reviewed articles, only one study (Kannampuzha & Hockerts, 2019) has a formative construct, a fact duly justified by the theoretical construction defined by the authors. In two articles, it is impossible to identify the construct's shape; in all 29 other studies, it was identified that the construct is reflective.

5.4 ANALYSIS TECHNIQUE

In the dimension of analysis technique, factor analysis in the development and scale validation studies is evaluated. Exploratory factor analysis (EFA) is used to identify complex interrelationships between variables without having any initial assumptions about





the factors. Confirmatory factor analysis (CFA) is an approach that allows verifying whether the variables are associated with their respective assumed factors based on a preexisting conceptual structure. The factorial analysis techniques allow the refinement and confirmation of the reliability and validity of the scales (Cabrera-Nguyen, 2010). Exploratory factor analysis is used to purify the measurement scale, while confirmatory factor analysis deals with scale validation (Worthington & Whittaker, 2006). Of the reviewed articles, one study did not explicitly use any factor analysis (Waltner, Riess, and Mischo, 2019). An article was developed using only confirmatory factor analysis. Another seven articles adopted only exploratory factor analysis as a refinement technique. The most frequent occurrence is integrating combined techniques (exploratory and confirmatory factor analysis) in the same study, as Netemeyer, Bearden, and Sharma (2003) recommended. In this condition, 21 studies demonstrated refinement through the concomitant use of the EFA and CFA techniques.

5.5 RESEARCH ORIENTATION AND PROTOCOL

The studies present diversity regarding adopting protocols that guide the methodological procedures used in conducting the analysis. It was possible to identify two patterns as to the orientation criterion and research protocol: i) a group of articles does not adopt any methodological protocol but prioritizes attention to the theoretical arguments of other studies in the construction of the scale, ii) the second group prioritizes adoption of the research protocol in the execution of the method. Studies that do not adopt a protocol in the execution of methodological procedures (six articles) pay more attention to theoretical articulation. However, they still use criteria and parameters such as, for example, Cronbach's alpha coefficient, customarily used to analyze the scale's internal consistency, as in the case of Das (2017). The other studies follow protocols widely used in the literature, which prescribe the decisive steps of their execution. The highest frequency is registered in the use of only one protocol (14 articles), followed by the integration of two protocols (in 7 articles) or three protocols (two articles).

The most widely used protocols are Churchill's (1991), which appears in eight studies; DeVellis' (2003) in six articles, and four other executions by Hair et al. (2009);





and two others adopting Hinkin protocols (1998), Worthington and Whittaker (2006). As different protocols are adopted, with different orientations regarding the theoretical content, the number of respondents and final items of the scales also varies considerably. Regarding the sample of respondents, the maximum frequency is the article by Dabija et al. (2017), with 1,752 valid responses. Regarding the number of items, after all the refinement processes, the lowest number of items is in the study by Larimian and Sadeghi (2019), with seven items, and the most significant amount is recorded in 64 items in the article by Soo Wee and Quazi (2005). Alpha coefficients vary between 0.55 (minimum value) and 0.97 (maximum value).

5.6 SCALE RANGE

The level of analysis of the scale's reach is related to the type of sample that was analyzed, considering the possibilities of having scales developed for the macro (country or regions), meso (production chains), or micro (companies, consumers) levels. The systematic review did not find scales applied at the macro level. Only in two studies was a meso scale developed, specifically addressing the development and validation of a scale on sustainable supply chain management practices. The other articles addressed research conducted at the micro level. They applied to i) the consumer context (nine articles), which deals with consumer behavior and intention, sustainable consumption awareness, and green consumption, and ii) the business context (14 studies), which explores performance, strategic orientation, sustainability, and entrepreneurship, marketing for sustainability, and corporate sustainability, among others. A third group of scales, applied at the micro level, addresses the context of universities (six articles). The breadth of the topics listed, from sustainable development to skills for sustainability, demonstrates that the scales of measurement in the broad field of sustainability address different topics within the broad field. They are not limited to related themes but also to different conceptual ends that are incorporated in discussions on sustainability.

5.7 RELIABILITY MEASURE





Reliability refers to the questions of the instruments' stability, consistency, and precision in light of the choice of statistical tests used to assess the reliability of scales according to the object to be measured (Bonett & Wright, 2014). In the reviewed studies, three modalities of reliability measurement were identified. First, the sum of the registered frequencies is higher than the total number of articles because some studies adopt more than one reliability criterion. The most commonly used measure is Cronbach's alpha (28 articles), which measures the degree of covariance between items on a scale, leading to the assessment of the scale's internal consistency. The smaller the sum of the variances of the items, the more consistent the scale will be (Allen et al., 2008). As supported by the literature, this is the most commonly used instrument for assessing reliability (Bonett & Wright, 2014); however, it is recommended to use composite reliability on scales with multiple factors. Within the perspective of the examiners' judgment—that is the analysis from the evaluators, which demonstrates the degree to which different examiners evaluate the same instrument—16 articles used this type of strategy. In a third aspect of the test-retest process, which consists of applying the same instrument in different situations, it was found that the procedure was used in eleven studies.

5.8 VALIDITY MEASURE

The level of validity measurement analysis addresses the techniques used to measure and improve the validity of the scales. Validity refers to the condition that the scale measures what is proposed (Borsboom, Mellenbergh, and van Heerden, 2004). As argued and prescribed by specialists, the validity and reliability measures are not independent of each other. Therefore, high reliability does not irreducibly guarantee the validity of the scale.

The most commonly used types of validity are discriminant (27 studies) and convergent (26 studies). Discriminant validity tests the hypothesis that the target measure is not unduly related to different constructs, with variables that should initially differ from each other. Convergent validity is obtained by correlating the focal instrument with another instrument that assesses a similar construct, expecting high correlations between the two. Both measures of validity are generally associated with other validity techniques. The third





most commonly used technique is content validity (14 articles). This measure tests the degree to which a scale includes all the items needed to represent the concept to be measured, and it can be assessed using a qualitative (committee of experts) or quantitative (content validity) approach. Nomological validity is tested in nine studies, which demonstrates how a scale correlates as theoretically expected with other measures. A tested theoretical model can lead to other deductions and inferences (Wanous, Reichers, and Hudy, 1997). Face and structural validity are found in 5-scale tests. Face validity consists of having specialists review the contents of the scale to assess whether they are appropriately appropriated in the way they are arranged. Structural validity tests whether a measure captures the hypothetical dimensionality of a construct (Wanous, Reichers, and Hudy, 1997).

Three of the reviewed scales used the criterion, concurrent and predictive validities. Criterion validity assesses when a result can be compared to a 'gold standard,' and predictive validity is assessed when the target test is first applied and then the 'gold standard.' Concurrent validity can be verified by applying both the target test and the 'gold standard' (Haynes, Richard, and Kubany, 1995). Two other validity tests were applied to only one article. The validity of known groups is performed when different individuals complete the research instrument, and then the groups' results are compared (Haynes, Richard, and Kubany, 1995). Cross-cultural validity demonstrates the extent to which the evidence supports the inference that the original and culturally adapted instruments are equivalent. Notably, the validity measures are not static properties; therefore, they vary according to the applicable circumstances, the sample analyzed, and the type and purpose with which the study was designed.

6 LIMITATIONS OF THE REVISED STUDIES

The development of each scale implies the generation of sets of limitations of the articles associated with different conditions. Then, based on the weaknesses of the studies, future studies can be drawn. The primary limitations noted in the studies are grouped into the following aspects:





- a) Method: i) limitations of the scale construction and validation method; ii) the data are extracted from self-reports and the respondents' perceptions.
- b) Sample: i) sample size; ii) characteristics of the subjects; iii) management of nonresponse rates; iv) heterogeneity of the samples; v) bias and response bias.
- c) Regionality: i) unique regional characteristics that make replicability impossible in other contexts; ii) difficulty in coding local elements in scales; iii) measurements used that are very or little sensitive to the context; iv) difficulty in measuring the influence of local factors on the results; v) imprecision of limits regarding spatial organization; vi) different cultural factors in the same region.
- d) Theoretical construct: i) specific focus on the dimensions that weaken the articulation of the construct; ii) weakness in the construction of the scale items; iii) fragility in the specifics of the relationships; iv) multiple mediating and moderating relationships in the same construct; v) very generalist or specific construct.
- e) Dynamics: i) dynamics of the topic addressed; ii) first attempt to build a scale on the theme; iii) difficulty in measuring dynamic dimensions.
- f) Field of activity: i) heterogeneous sectors covered on the same scale; ii) impossibility of replicating the scale for different branches of the same sector; iii) sector specificities cannot be captured by general scales.

The limitations associated with the method and sample listed are common to research with a quantitative approach. The aspect of regionality, theoretical construct, the dynamism of the topic addressed, and the branch of activity is inherent to the context and development specificities of each scale.

7 DISCUSSION OF RESULTS





The analyzed articles dealing with scale construction and validation show that the panorama of research in sustainability is based on a plurality of themes. The greatest concentration is on studies on sustainable behavior and consumption, as well as on environmental strategies and management. The area of behavior and consumption includes topics such as green consumption (Ahmad, Omar, and Quoquab, 2019), environmentally oriented consumption (Kassing et al., 2010), and awareness of conscious consumption (Yan and She, 2011). The strategy field addresses issues such as critical factors for environmental management (Gupta & Agrawal, 2017) and strategic guidance for marketing (Meiboudi et al., 2017).

The studies are concentrated in contexts that could be categorized as predominant in developed or emerging countries. As the measures of reliability and validity are associated, among other factors, with the context where the instruments are applied and tested, it is postulated, in this sense, that there is a lack regarding the measurement of these same phenomena in different contexts (that is, a critical element to assure the measure's invariance assumption). For example, it would be plausible to assume that cross-cultural factors will condition the impossibility of faithfully applying the same scale to another sociocultural context. Regarding regionality, studies are concentrated in Asia and Europe. Scales applied to continents such as Central or South America or Africa were not found, which indicates that the theme can be expanded for evaluation in countries on those continents.

The emphasis given by the instruments is centered on manufacturing companies and universities. Significant and essential sectors for sustainability, such as services, are neglected. Several opportunities could be developed if this and other sectors yet to be explored were considered to construct future scales. In this same sense, it is registered that the dimension of analysis is restricted to the micro level. Supply chains, for example, are little explored and offer a wealth of contributions to understanding the broad topic in question. The same situation is perceived at the macro level due to the absence of scales.

The techniques for measuring the reliability and validity of the instruments are diversified in terms of the number of tests and the number of executions for purification and refinement. For example, the scales proposed by Kannampuzha and Hockerts (2019) and Haan et al. (2018) recorded only one type of reliability test. Other scales, such as





Wang (2015), tested reliability more widely. Robertson and Barling (2017) performed six different tests for the validity measures. For example, authors like Papadas, Avlonitis, and Carrigan (2017) used five other validity verification techniques. As for the number of times, the questionnaires were applied, between the initial generation of the scale and the final presentation of the scale, there is a minimum variation of one application to a maximum variation of six applications. The continuous repetition of the application of the instrument gives the character of purification and refinement of the items on the scale. In the case of six repetitions (Robertson & Barling, 2017), the steps of i) evaluation by expert/expert, ii) test, iii) purification and refinement, and iv) validation were performed repeatedly six times. Another relevant aspect to be highlighted is that in seven studies, it was impossible to identify the use of an execution protocol prescribed in the literature. The construction and validation stages were developed according to the authors' guidelines, which may compromise the replicability of the research protocol when it needs to be clarified in the body of the article itself.

The themes on which the scales were developed demand special attention. Several fields and research areas, especially those related to more contemporary and dynamic themes, such as the circular economy, must be validated. The most promising advance, in this sense, is in the direction of building indicators to measure the phenomenon in specific sectors or contexts.

Notably, the study by Corona et al. (2019) mapped roundness metrics applied to products and services and identified seven measurement indices, nine evaluation indicators, and three evaluation structures. However, studies deal with only some of the circular economy. Kristensen and Mosgaard (2019) analyzed 30 indicators at the micro level, and they identified that there is no standard way of measuring the topic in general. The study by Secco et al. (2020) proposes a model to measure the circular economy in the pig industry. The proposal contemplates the environmental, economic, and social aspects of this production chain and assists decision-making regarding the interaction of the circular economy with waste management. Linder et al. (2020) apply a metric developed from indicators at the product level to assess the relationship between the circularity of materials and the environmental impact. From a macroenvironmental perspective, Silvestri, Spigarelli, and Tassinari (2020) construct two indicators for static





and dynamic evaluation of the performance obtained regarding applying the premises of the circular economy in the specific context of the European Union.

In this way, there is a significant opportunity to contribute to sustainability studies, especially in the circular economy context, to promote progress in measuring the circularity of resources. New scales are needed and will be more solid if one considers the limitations and weaknesses of those elaborated by previous studies.

8 PROPOSED AGENDA FOR FUTURE RESEARCH

Based on the weaknesses noted opportunities for future studies are presented. Theoretical propositions are presented for validation in future studies. It is essential to highlight that the opportunities are aligned not only with theoretical aspects but also with the context, research trends, and themes that can promote progress toward sustainability through the development of measurement scales.

Proposition 1: A delimited and clearly defined (specific) theoretical construct positively influences the construction and validation of effective measurement scales.

The scale construction process begins with defining the theoretical construct that guides its development. With a precise definition of what will be effectively measured, the instrument will be accurate and correspond to the correct assessment of the measured phenomenon. Furthermore, constructing the theoretical model delimits how the measured object will be articulated with the propositions identified in the literature. This care removes subjectivity, dubiousness, and ambiguity and enables epistemological rigor for later applications (DeVellis, 2003). The concern about the reliability and validity of scales is addressed by Turner and Zolin (2012); they discourage the free elaboration of scales by adopting a research protocol with a well-defined theoretical construct.

The way of approaching the theoretical construct involves the following research possibilities: (i) Develop more comprehensive theoretical frameworks on the topic; (ii) Develop a more in-depth systematic review on the topic to capture broadly related practices; (iii) Evaluate and survey the antecedents and strength of relationships in the





constructs; (iv) Test existing scales with other theoretically related constructions; and (v) Include items transversal to the scales.

Proposition 2: The sample size influences the construction and validation of a practical measurement scale.

Applying the same research instrument to different samples can result in significantly different values (Maroco, Maroco, and Campos, 2014). Therefore, sample homogeneity can represent a critical factor for the instrument's reliability because variability is highly desired for the purification and refinement of the measurement instruments. Furthermore, different populations manifest different preferences and behaviors, especially concerning cultural factors. Therefore, the same measurement instrument applied in different contexts, with more or less homogeneous samples, leads to different reliability indexes.

From the perspective of the sample criterion, the following research opportunities are launched: (i) Explore the use of the scale in different sociodemographic contexts; (ii) Test scales in other locations, branches of activity, and different audiences; (iii) Expand the sample size for validation; (iv) Expand the branches of activities covered by the scale; (v) Explore higher levels of reach of research subjects beyond the company level; (vi) Address the sector supply chain under analysis; and (vii) Include context-specific items.

Proposition 3: The methodological process used in the construction and scale validation influences the instrument's effectiveness.

The construction of measurement scales requires the development of an instrument that associates qualitative concepts with quantitative parameters and metrics (Pooja & Sagar, 2012). Therefore, the methodological process that involves the creation of the scale must describe which rules were used or which protocol was followed so that the study's replication is guaranteed. Furthermore, using integrated and mixed methods brings the researcher closer to a more assertive answer about the studied phenomenon. With this, the researcher can direct more significant effort toward the theoretical articulation that guides the study.

For the method improvement criterion, the following improvement proposals are defined: (i) use integrated and mixed methods to increase the external validity of scales; (ii) expand the use of qualitative methods for the initial construction of scales; (iv) use





more refined scale refinement techniques; (v) validate scale items; (vi) check the influence of sociodemographic factors on the strength of relationships; and (vii) integrate different validity and reliability methods.

Proposition 4: Emerging themes positively influence the construction and validation of effective scales.

The broad field of knowledge on sustainability articulates other areas that cut across business practices, consumer behavior, and the economic system. Above all, advancing knowledge and practice follow several routes in times of crisis and highly volatile and dynamic scenarios. In transitioning to a sustainable society, the circular economy is the new sustainability paradigm (Geissdoerfer et al., 2017). Quantifying the circularity of products and services is essential in defining policies and strategies for companies to provide alternatives developed from facts (Corona et al., 2019). The development of measurement scales is the most appropriate way for researchers to map aspects associated primarily at the micro level (companies and consumers). These measures are developed based on indicators. However, robust and consistent works on construction and scale validation have yet to be identified to discuss the circular economy. Given the recent global crisis events, other related topics could also be discussed.

Regarding updating and incorporating new themes, the following possibilities are listed: (i) Develop and validate scale on the theme of the circular economy; (ii) Develop and validate sustainability scale in crisis scenarios; and (iii) Develop scale on sustainable consumption behavior in crisis scenarios.

Studies can be developed from the proposals for future research to fill the theoretical gaps and advance the field. The items listed can be adapted to specific sectors or contexts to align research in the field of sustainability with assumptions widely disseminated in more stable areas in the development of scales. Under a more in-depth analysis of the sustainability area, the scales can measure this association with other related areas, which have a relationship but still need to be explored by applying their method and protocol.

9 FINAL CONSIDERATIONS AND LIMITATIONS OF THE RESEARCH





The present study aims to analyze the specificities of measurement scales with a scope of sustainability; analytical categories were drawn up to evaluate the articles selected in the review. From the units of analysis, parameters on successive repetitions of elements in the studies were defined and related to the dimensions analyzed. The analysis criteria were defined from derivations of scale development paradigms prescribed by Churchill (1991), DeVellis (2003), Hinkin (1998), Netemeyer, Bearden, and Sharma (2003), and Hair et al. (2009). The stages listed in the scale development protocols prescribed by the authors were articulated and formed a panel through which the characteristics of the studies were extracted. In addition to the aspects already discussed, related to the analytical nature, the scope of the scale, and measures of reliability and validity, nine stages common to the protocols were defined and could be observed in the development and refinement of the scales.

According to the protocols listed and the analysis of the revised articles, the construction and validation of scales in sustainability go through the following stages:

- i) Articulation, definition, and a theoretical domain: This stage is adopted by all the reviewed studies and is prescribed as indispensable in developing scales. The stage confers reliability to the instrument by enabling the theoretical construct to represent what it proposes to measure.
- ii) Qualitative stage: In general, integrating the qualitative stage with the construction of a scale allows the approach of the phenomenon to be analyzed by different methods and helps in the generation of the initial scale. Twelve articles follow this stage.
- iii) Generation of the initial scale: The initial presentation occurs before applying the purification and refinement tests (stage adopted by all authors).
- iv) Evaluation by an expert consists of placing the initial scale in a specialist evaluation to verify consistency and theoretical relevance (adopted in 17 studies).
- v) Test: Generation of the initial test prior to the final presentation of the scale (performed on all articles).





- vi) Purification and refinement: application of reliability and validation techniques to refine the scale (performed on all).
- vii) Validation: stage performed in conjunction with the previous stage and performed unanimously.
- viii) Revalidation (repetition of steps 5 to 7): the process of repeated executions of the stages of testing, purification, refinement, and validation is adopted at 14 scales. The result is the presentation of a leaner scale with greater validity and reliability. The process is repeated two to six times, as each article reports.
- ix) Presentation of the final scale: At the end of creating and improving the scales, the proposal with the final items of the instrument is presented.

Authors' efforts to check the rigor and relevance of theoretical constructs, as well as the reliability and validity of the scales applied to the measurement of phenomena associated with sustainability, can be amplified by the adoption of a protocol for scale development. However, as recorded by Straub (1989, 2004), it is worrisome that research does not follow any other protocol for elaborating their scales and does not adopt a protocol as a standard reference.

The main theoretical contribution of the article is to diagnose the state of the creation of scales in sustainability and to provide subsidies for future studies. For some subareas, research with a scale creation character must be more present in this field. The theoretical contribution is to diagnose the state of the creation of scales in sustainability and to provide subsidies for future studies. The studies were revised for analytical aspects and to answer the following research questions:

1. What are sustainability measurement scales' specific limitations and challenges?
2. How can the findings of this study be applied in practical settings, such as in the development of sustainability management policies and strategies?
3. Are there any current efforts to standardize sustainability measurement scales, and if so, what are they? The scope of scales, validity and reliability techniques, and distinct creation, refinement, and purification stages were defined.





Several possible paths are proposed for future research addressing the topic. New and highly dynamic fields, such as the circular economy, still need to be explored by this research approach and demand the attention of researchers. The practical contribution is to signal opportunities for companies, universities, and the government regarding the possibilities to be explored and stimulated to develop knowledge and practice about the field.

The study's limitations are limited to aspects related to content and method. The limitations on content are related to the identification of a small number of articles included in the review. The sustainability theme is still consolidated, affecting the number of scales in the academic literature. The need for more previous studies reduces the breadth of results. For the method, the analysis and coding in systematic reviews involve the authors' capacity for exploration and cognition. The results may be limited, given this condition. The choice of a restricted number of keywords is a limitation that restricts the scope of the study. The procedure was adopted to limit the selection of studies with a specific scope in the subject under analysis. The absence of subcategories and subcodes to deepen the analysis also limits the study regarding the exhaustion of information extraction. In addition to the recommendations already listed, it is recommended that future research proceed with the analysis of subcategories based on the dimensions listed.

1. What are some specific examples of sustainability measurement scales, and how are they used in practice?
2. How can the findings of this study be applied in real-world contexts, such as in business or policy-making?
3. Are there any limitations or potential drawbacks to using sustainability measurement scales, and if so, what are they?

REFERENCES

Adams, R., Jeanrenaud, S., Bessant, J., Denyer, D., Overy, P., 2016. Sustainability-oriented Innovation: A Systematic Review. *International Journal of Management Reviews*, 18(2), 180–205. doi:10.1111/ijmr.12068





Ahmad, S., Omar, R., Quoquab, F., 2019. Corporate Sustainable Longevity: Scale Development and Validation. SAGE Open, 9(1), 215824401882237. doi:10.1177/2158244018822379

Alvarado-Herrera, A., Bigne, E., Aldas-Manzano, J., Curras-Perez, R., 2017. A Scale for Measuring Consumer Perceptions of Corporate Social Responsibility Following the Sustainable Development Paradigm. Journal of Business Ethics, 140(2), 243–262. doi:10.1007/s10551-015-2654-9

Asokan, V. A., Yarime, M., Onuki, M., 2019. A review of data-intensive approaches for sustainability: methodology, epistemology, normativity, and ontology. Sustainability Science. doi:10.1007/s11625-019-00759-9

Biasutti, M., Frate, S., 2016. A validity and reliability study of the Attitudes toward Sustainable Development scale. Environmental Education Research, 23(2), 214–230. doi:10.1080/13504622.2016.1146660

Bonett, D. G., Wright, T. A., 2014. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. Journal of Organizational Behavior, 36(1), 3–15. doi:10.1002/job.1960

Bonisoli, L., Galdeano-Gómez, E., Piedra-Muñoz, L., 2018. Deconstructing criteria and assessment tools to build agri-sustainability indicators and support farmers' decision-making process. Journal of Cleaner Production, 182, 1080-1094, doi: 10.1016/j.jclepro.2018.02.055.

Borsboom, D., Mellenbergh, G. J., van Heerden, J., 2004. The concept of validity. Psychological Review, 111(4), 1061- 1071.

Brink, M., Hengeveld, G. M., Tobi, H., 2020. Interdisciplinary measurement: A systematic review of the case of sustainability. Ecological Indicators, 112, 106145. doi:10.1016/j.ecolind.2020.106145

Cabrera-Nguyen, P., 2010. Author Guidelines for Reporting Scale Development and Validation Results in the Journal of the Society for Social Work and Research. Journal of the Society for Social Work and Research, 1(2), 99–103. doi:10.5243/jsswr.2010.8

Carson, R., 1962. Silent Spring. Fawcett Publications, Greenwich.

Churchill, G.A., 1991. Marketing Research: Methodological Foundation, 5th Edition, The Dryden Press, Fort Worth.

Coltman, T., Devinney, T. M., Midgley, D. F., Venaik, S., 2008. Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61(12), 1250–1262. doi:10.1016/j.jbusres.2008.01.013





Comin, L. C.; Aguiar, C. C.; Sehnem, S.; Yusliza, M.Y.; Cazella, C. F.; Julkovski, D. J., 2019. Sustainable business models: a literature review. *Benchmarking: An International Journal*, (), BIJ-12-2018-0384–. doi:10.1108/BIJ-12-2018-0384

Corona, B., Shen, L., Reike, D., Rosales Carreón, J., Worrell, E., 2019. Towards sustainable development through the circular economy—A review and critical assessment on current circularity metrics. *Resources, Conservation and Recycling*, 151, 104498. doi:10.1016/j.resconrec.2019.104498

Dabija, D.-C., Postelnicu, C., Dinu, V., Mihăilă, A., 2017. Stakeholders' perception of sustainability orientation within a major Romanian University. *International Journal of Sustainability in Higher Education*, 18(4), 533–553. doi:10.1108/ijsh-10-2015-0169

Devellis, R. F., 2003. *Scale Development: Theory and Applications*. 2nd ed. Thousand Oaks, CA: Sage.

Elkington, J., 1997. *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Capstone, Oxford.

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., Hultink, E. J., 2017. The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*, 143, 757–768. doi:10.1016/j.jclepro.2016.12.048

Gupta, S., Agrawal, R., 2017. Environmentally Responsible Consumption: Construct Definition, Scale Development, and Validation. *Corporate Social Responsibility and Environmental Management*, 25(4), 523–536. doi:10.1002/csr.1476

Haan, M., Konijn, E. A., Burgers, C., Eden, A., Brugman, B. C., Verheggen, P. P., 2018. Identifying Sustainable Population Segments Using a Multi-Domain Questionnaire. *Social Marketing Quarterly*, 152450041879401. doi:10.1177/1524500418794019

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2009. *Multivariate Data Analysis*, Pearson- Prentice Hall, Englewood Cliffs, NJ.

Haynes, S. N., Richard, D. C. S., Kubany, E. S., 1995. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3). 238-247.

Hays, J., Reinders, H., 2020. Sustainable learning and education: A curriculum for the future. *International Review of Education*. doi:10.1007/s11159-020-09820-7

Hinkin, T. R., 1998. A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, 1(1), 104–121. doi:10.1177/109442819800100106

Kannampuzha, M., Hockerts, K., 2019. Organizational social entrepreneurship: scale development and validation. *Social Enterprise Journal*. doi:10.1108/sej-06-2018-0047





Kapitan, S., Kennedy, A.-M., Berth, N., 2018. Sustainably superior versus greenwasher: A scale measure of B2B sustainability positioning. *Industrial Marketing Management*. doi:10.1016/j.indmarman.2018.08.003

Khan, E. A., Quaddus, M., 2015. Development and Validation of a Scale for Measuring Sustainability Factors of Informal Microenterprises – A Qualitative and Quantitative Approach. *Entrepreneurship Research Journal*, 5(4). doi:10.1515/erj-2014-0017

Kristensen, H. S., Mosgaard, M. A., 2019. A review of micro level indicators for a circular economy – moving away from the three dimensions of sustainability? *Journal of Cleaner Production*, 118531. doi:10.1016/j.jclepro.2019.118531

Larimian, T., Sadeghi, A., 2019. Measuring urban social sustainability: Scale development and validation. *Environment and Planning B: Urban Analytics and City Science*, 239980831988295. doi:10.1177/2399808319882950

Likert, R., 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 22
140, 55.

Linder, M., Boyer, R. H. W., Dahllöf, L., Vanacore, E., Hunka, A., 2020. Product-level inherent circularity and its relationship to environmental impact. *Journal of Cleaner Production*, 121096. doi:10.1016/j.jclepro.2020.121096

Mani, V., Agarwal, R., Gunasekaran, A., Papadopoulos, T., Dubey, R., Childe, S. J., 2016. Social sustainability in the supply chain: Construct development and measurement validation. *Ecological Indicators*, 71, 270–279. doi:10.1016/j.ecolind.2016.07.007

Maroco, J., Maroco, A., Campos, J., 2014. Student's Academic Efficacy or Inefficacy? An Example on How to Evaluate the Psychometric Properties of a Measuring Instrument and Evaluate the Effects of Item Wording. *Open Journal of Statistics*, 4, 484-493. doi: 10.4236/ojs.2014.46046.

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. 1972. *The limits to growth*. New York, 102, 27.

Meiboudi, H., Lahijanian, A., Shobeiri, S. M., Jozi, S. A., Azizinezhad, R., 2017. Development and validation of sustainability criteria of administrative green schools in Iran. *Journal of Environmental Management*, 197, 605–609. doi:10.1016/j.jenvman.2017.04.045

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., ... Stewart, L. A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). doi:10.1186/2046-4053-4-1

Netemeyer, R. G., Bearden, W. O., Sharma, S., 2003. *Scaling procedures: Issues and applications*. Sage Publications.





Nilashi, M., Rupani, P. F., Rupani, M. M., Kamyab, H., Shao, W., Ahmdi, H., ... Aljojo, N., 2019. Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach. *Journal of Cleaner Production*, 118162. doi:10.1016/j.jclepro.2019.118162

Papadas, K.-K., Avlonitis, G. J., Carrigan, M., 2017. Green marketing orientation: Conceptualization, scale development and validation. *Journal of Business Research*, 80, 236–246. doi:10.1016/j.jbusres.2017.05.024

Pizzi, S., Caputo, A., Corvino, A., Venturelli, A., 2020. Management research and the UN sustainable development goals (SDGs): A bibliometric investigation and systematic review. *Journal of Cleaner Production*, 276(), 124033–. doi:10.1016/j.jclepro.2020.124033

Pooja, S., Sagar, M., 2012. High impact scales in marketing: a mathematical equation for evaluating the impact of popular scales. *Advances in Management*, 5(4), 31-48.

Robertson, J. L., Barling, J., 2017. Toward a new measure of organizational environmental citizenship behavior. *Journal of Business Research*, 75, 57–66. doi:10.1016/j.jbusres.2017.02.007

Roca, L. C., Searcy, C., 2012. An analysis of indicators disclosed in corporate sustainability reports. *Journal of cleaner production*, 20(1), 103-118. doi: 10.1016/j.jclepro.2011.08.002

Secco, C., Mendes da Luz, L., Pinheiro, E., Puglieri, F. N., Piekarski, C. M., Carlos de Francisco, A., Cereja Seixas Freire, F. M., 2020. Circular economy in the pig farming chain: Proposing a model for measurement. *Journal of Cleaner Production*, 121003. doi:10.1016/j.jclepro.2020.121003

Sehnm, S., Vazquez-Brust, D., Pereira, S. C. F., Campos, L. M., 2019. Circular economy: benefits, impacts and overlapping. *Supply Chain Management: An International Journal*. doi: 10.1108/SCM-06-2018-0213

Silvestri, F., Spigarelli, F., Tassinari, M., 2020. Regional development of Circular Economy in the European Union: a multidimensional analysis. *Journal of Cleaner Production*, 120218. doi:10.1016/j.jclepro.2020.120218

Soo Wee, Y., Quazi, H. A., 2005. Development and validation of critical factors of environmental management. *Industrial Management & Data Systems*, 105(1), 96–114. doi:10.1108/02635570510575216

Stevens, S., 1946. On the Theory of Scales of Measurement. *Science*, 103(2684), 677-680.

Straub, D., 1989. Validating Instruments in MIS Research. *MIS Quarterly*, 13(2), 147. doi:10.2307/248922





Straub, D., Gefen, D., 2004. Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, 13. doi:10.17705/1cais.01324

Taecharungroj, V., Tachapattaworakul Suksaroj, T., Rattanapan, C., 2018. The place sustainability scale: measuring residents' perceptions of the sustainability of a town. *Journal of Place Management and Development*. doi:10.1108/jpmd-04-2017-0037

Tranfield, D., Denyer, D., Smart, P., 2003. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207–222. doi:10.1111/1467-8551.00375

Trompenaars, F., Hampden-Turner, C., 1998. *Riding the Waves of Culture*. New York: McGraw-Hill. UNDP. 2001. "Human Development Reports."

Turner, R., Zolin, R., 2012. Forecasting Success on Large Projects: Developing Reliable Scales to Predict Multiple Perspectives by Multiple Stakeholders over Multiple Time Frames. *Project Management Journal*, 43(5), 87–99. doi:10.1002/pmj.21289

Waltner, E.-M., Riess, W., Mischo, C., 2019. Development and Validation of an Instrument for Measuring Student Sustainability Competencies. *Sustainability*, 11(6), 1717. doi:10.3390/su11061717

Wang, Y.-F., 2015. Development and Validation of the Green Food and Beverage Literacy Scale. *Asia Pacific Journal of Tourism Research*, 21(1), 20–56. doi:10.1080/10941665.2015.1016050

Wanous, J. P., Reichers, A. E., Hudy, M. J., 1997. Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, 82(2), 247–252. doi:10.1037/0021-9010.82.2.247

WCED, 1987. *World Commission on Environment and Development- Our Common Future*. United Nations.

Worthington, R. L., Whittaker, T. A., 2006. Scale Development Research. *The Counseling Psychologist*, 34(6), 806–838. doi:10.1177/0011000006288127

Yan, J., She, Q., 2011. Developing a Trichotomy Model to Measure Socially Responsible Behaviour in China. *International Journal of Market Research*, 53(2), 253–274. doi:10.2501/ijmr-53-2-253-274

