Global Insights on Scale Development Practice: Shaping Tourism Research

Global Insights on Scale Development Practice: Shaping Tourism Research

Introduction

Measurement scales provide destination managers with valuable insights into various aspects of their destination, enabling data-driven decision-making and strategic planning. Thus, developing reliable and valid measurement scales is essential for obtaining accurate data and insights in tourism research. The substandard quality of certain scholarly works has often been ascribed to the inadequate measurement researchers employ to evaluate their chosen variables of interest. Consequently, some scholars have proposed a process for creating measures with desirable psychometric properties (Churchill, 1979; DeVellis, 2003; Hinkin et al., 1997). Generally, the scale development process can be delineated into three basic steps: item generation, theoretical analysis, and evaluation of psychometric properties (Morgado et al., 2017).

Over the past decades, hundreds of scales – either originally developed or adapted from other disciplines—have been used to assess the abstract concepts of various stakeholders in hospitality and tourism management. Scales have primarily been developed, with varying contexts and abstraction, for the tourism experience and behavior, tourism operators, and residents. (Gursoy et al., 2014) Such knowledge allows a deeper explanation and prediction of constructs relevant to tourists, residents, and service providers. The insights from these scales inform tourism operators and marketers, often from the tourist's perspective and sometimes at the destination level. Insights also inform branded service design and execution, sustainability, and strategy for tourism development, including with respect to local residents.

Despite the plethora of scales developed and validated in this field, a systematic review of these scales is notably lacking. Scale instruments contribute to assessing the hypothesized relationships with other constructs or behaviors. A comprehensive understanding of the existing measurement scales within the tourism domain is crucial. A few researchers have analyzed existing scales from the perspectives of scope and methodology; however, their focus has primarily been on the topics and scopes of the scales, or the psychometric properties of the scales (Koc & Ayyildiz, 2022). There remains a significant opportunity to examine the theoretical and methodological rigor of the existing scales from the perspective of the whole development process. To address the gaps mentioned above, the current study reviewed scale development studies in the tourism management field, to obtain bibliometric information on existing scales, thereby enhancing understanding of their status, and providing suggestions for future scale development research.

Method

The current study followed the rigorous and comprehensive data collection process of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISM) method to collect relevant articles (Moher et al., 2009). The PRISM consists of four stages: identification, screening, eligibility, and included.

Utilizing the Web of Science and EBSCO databases, the study ensured the inclusion of peer-reviewed and validated studies. A combination of specific keywords was employed to identify relevant articles in the field of hospitality and tourism management (HTM). The search query was "scale development AND hospitality OR tourism OR travel," targeting journal articles' titles, abstracts, and keywords. While the search was restricted to articles published in English to maintain consistency, it imposed no temporal restrictions, allowing for a comprehensive review of studies from past to present. The initial search in the Web of Science yielded 98 articles, and EBSCO returned 70 articles (retrieved on April 1st, 2023). In total, 168 articles were subjected to a rigorous screening process following the PRISMA flowchart to ensure their quality and pertinence.

Upon the elimination of 23 duplicate articles, the remaining 145 articles were scrutinized by examining their titles, abstracts, and keywords. This examination led to the exclusion of records that diverged from the core focus of scale development within the hospitality and tourism management field. As a result, an additional 7 papers were removed from consideration. During the full-text assessment phase, 3 articles were further excluded for various reasons, such as not being an empirical scale development study, or solely addressing the cross-cultural validation of an existing scale. Consequently, the remaining 135 articles met the inclusion criteria and were selected for the final analysis and synthesis (Figure 1).

The coding procedure encompassed two distinct phases: an initial round to establish concordance among coders, followed by a second round of formal coding. In the preliminary round, four researchers independently coded a random sample of 23 articles from the dataset. Cohen's Kappa (*k*) statistic was employed to calculate the inter-rater reliability of the indicator Predictive validity (Haley & Osberg, 1989), yielding an acceptable agreement level (*k*=0.65)(Landis & Koch, 1977). Discrepancies were addressed through meticulous review and reached consensus among the researchers. After this, the second phase entailed the formal coding of the remaining 112 articles. The study further assessed the consistency of coding across various sets of articles by the four researchers. Specifically, the employment of confirmatory factor analysis (CFA)—an essential phase in the scale development process—was examined. The frequency of CFA use in the articles, as coded by the researchers, uniformly ranged from 90.0% to 95.2% (Coder1=95.2%, Coder2=92.3%, Coder3=90.0%, Coder4=92.9%), thus substantiating the inter-rater reliability.

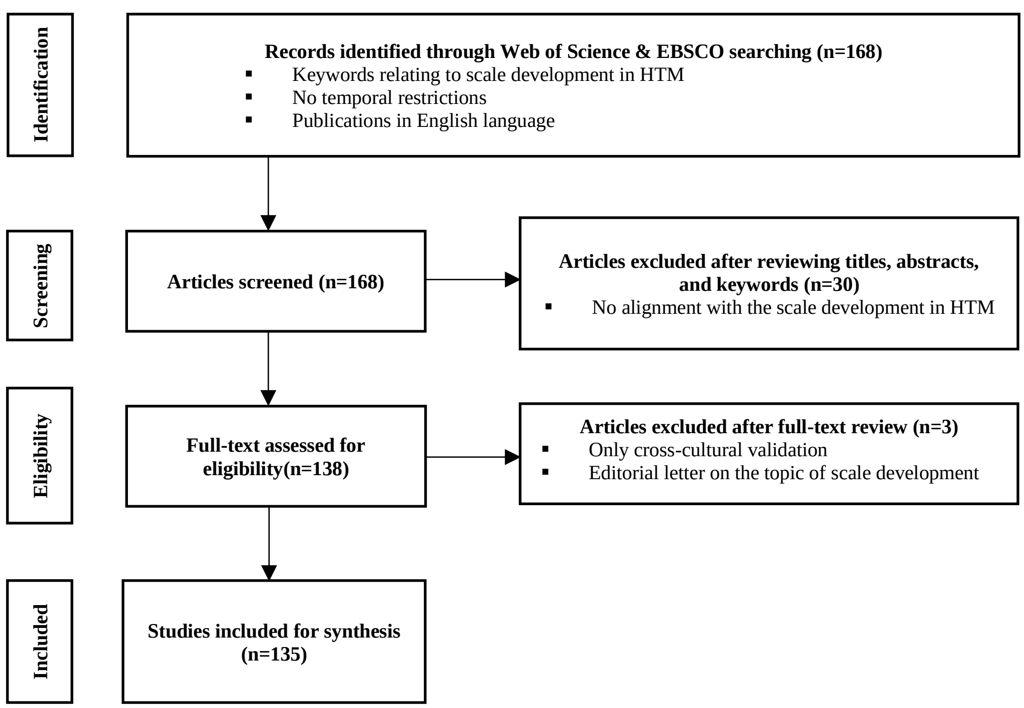

Figure 1. PRISMA Flow Diagram of Systematic Article Research and Selection

Bibliometric Information on Existing HTM Scales

The current study conducted a review of 135 studies published in the period from 1996 to 2023, and the current study followed a structure similar to that of a previous study by Morgado et al. (2017) when reporting results. A majority (77%, n=104) were published after 2016, reflecting a growing interest and scholarly focus in this area. Remarkably, only a single study dated back to the last century. Regarding the design of the scale development process, 78 studies adhered to the scale development process proposed by Churchill (1979). Additionally, 28 studies conformed to the guidelines proposed by DeVellis and Thorpe (2021), while two studies followed Menor and Roth (2007). Furthermore, seven studies utilized the framework established by Hinkin (1995), and three studies implemented the methodology described by Hinkin et al. (1997). It is noteworthy that most of the articles did not strictly follow a single framework but rather integrated multiple guidelines in their scale development process.

Step 1—Item Generation Most (77.8%, n=105) of the articles applied both deductive and inductive strategies for item generation. A smaller proportion, 17% (n=23), employed solely deductive methods, and 5.2% (n=7) relied exclusively on inductive methods. For deductive approaches within item generation, a substantial majority (94.8%, n=128) utilized literature reviews, and 59.3% (n=80) incorporated existing scales. Regarding inductive methodologies, interviews were conducted in 58.5% (n=79) of the studies, while focus groups were utilized for item generation in 36.3% (n=49) of the cases.

Step 2—Theoretical Analysis In the theoretical analysis phase, a predominant share of studies (90.4%, n=122) employed reflective methods for scale development. A smaller fraction (4.4%, n=6) utilized formative methods, and 5.2% (n=7) implemented a combination of both reflective and formative approaches. To refine items, most studies solicited the opinions of experts (91.9%, n=124), and over half (54.1%, n=73) considered the opinions of the target population. Moreover, 40.7% (n=55) of the studies opted to use just one of these methods (either expert or population judges).

Step 3 – Psychometric Analysis In evaluating construct validity, the predominant analytical methods reported were Exploratory Factor Analysis (EFA) at 94.1% (n=127), CFA at 92.6% (n=125), convergent validity at 94.8% (n=128), and discriminant validity at 93.3% (n=126). A substantial majority of studies employed both EFA and CFA (86.7%, n=117). Only two studies (1.5%) did not utilize factor analysis in their scale validation phase. Additionally, 56.3% (n=76) of studies verified nomological validity, 70.4% (n=95) assessed criterion validity, and 46.7% (n=63) examined internal validity.

Regarding the reliability of the studies, all investigated internal consistency reliability. The item-total correlation reliability was the second most employed technique, utilized in 86.7%, (n=117) of the studies. Additionally, 28.1% (n=38) of the articles addressing scale development took common method bias into account. A smaller proportion, 19.3% (n=26), applied split-half sampling methods. Pretesting was conducted in 68.9% (n=93) of the studies, and pilot testing was carried out in 74.8% (n=101) of the studies before executing the formal survey.

Sample size and software in step3 and number of items In step 3, concerning sample size, a significant 79.3% (n=107) of studies adhered to a sample size exceeding the commonly accepted guideline of at least 10 participants for each scale item (Schwab, 1980). Regarding item reduction, nearly half of the studies (49.6%, n=67) reported a reduction greater than 50% of the initial item pool through the validation process, while a small percentage (5.9%, n=8) did not explicitly report the initial number of items.

For the data analysis software utilized, Amos was employed in 37% (n=50) of the studies, LISREL was used in 15.6% (n=21), and various other software tools (e.g., R, Python) were reported in 17% (n=23) of the studies. Notably, 17.8% (n=24) of the studies did not disclose the software used.

Most concerned topics in the existing literature The authors further summarize the research topics that have been examined in the existing literature. As shown in Table 1, the existing HTM scales were developed from three perspectives: tourists, residents, and organizations. The most concerned topics focus on tourists' behavior, experience, perception, motivations, and attitudes, as well as their specific characteristics. For instance, the interactions, involvement, and attachment of tourists have garnered significant attention, alongside the co-creation of tourist experiences, and their satisfaction within context-specific scenarios like gastronomy or medical tourism. Tourist perceptions and attitudes have also been explored, especially regarding technology and risk, with a singular study looking at backpacker personal development.

In the realm of residents, the perception of tourism impacts and the empowerment it can provide are prominent subjects of interest, along with attitudes towards sustainable and gambling tourism. As for organizations, the literature pays attention to emotional labor, leadership, corporate social responsibility (CSR), and organizational culture. Service management is another major area, with studies frequently addressing service quality, destination image, and tourist constraints. This comprehensive examination from multiple perspectives underscores the multifaceted nature of HTM research and its scales, reflecting the complex interplay between individual behaviors, perceptions, community attitudes, and organizational practices in the tourism and hospitality sector.

Table 1. Most Concerned Topics in the Existing HTM Scales.

Perspective	Research topics	Examples of most concerned research topics	Number
Tourists	Tourists’ behavior	Interactions, Involvement, Attachment	27
	Tourist’ motivation	Casino gaming motivation, Cruise motivation, Food tourist motivation	12
	Tourists’ experience	Co-creation, Memorable experience, Satisfaction, Contexts-specific (e.g., Gastronomy, Medical tourism)	12
	Tourists’ perception	Trust, Risk perception, Perceived value	15
	Tourists’ Attitudes	Attitudes toward technology, risk	5
	Tourists’ characteristics	Backpacker personal development	1
Residents	Residents’ perception	Perceived tourism impacts, Empowerment	6
	Residents’ Attitude	Attitudes toward sustainable/gambling tourism	3

Organizations	Organizational behavior	Emotional labor, Leadership, CSR, Organizational culture	18
	Service management	Service quality, Destination image, Tourist constraints	36

Suggestions to improve scale development efforts

To enhance the efforts in scale development, a multifaceted approach is advocated, integrating both methodological rigor and theoretical depth. Suggestions have been proposed following the three basic steps: item generation, theoretical analysis, and psychometric properties.

Item generation. The majority of studies combined deductive and inductive strategies to generate an initial pool of items, aligning with the recommendations by DeVellis (2003). The rationale behind this approach is to establish a rich source from which a scale can emerge; typically, a larger pool of items is preferable. However, researchers must remain vigilant about content redundancy as it also correlates to internal consistency. Beyond essential literature reviews, the integration of innovative qualitative methods, such as thematic analysis of travel reviews, is useful in revealing new dimensions of the construct (Lin et al., 2022).

Theoretical analysis. The employment of either reflective, formative, or combination methods should be justified. As noted by Coltman et al. (2008), the selection of a measurement model— be it reflective, where measures are outcomes of unobserved latent variables, or formative, where measures act as causes of constructs (Edwards, 2011)— should be justified both theoretically and empirically. This justification needs to consider the construct's nature, the causality direction between indicators and the latent construct, and the indicators' characteristics. The validity of the scale and the structural relationships with other constructs would be influenced by the adoption of an unsuitable measurement model, which in turn, could diminish the applicability of theories.

Psychometric properties. The employment of psychometric properties is evident in the scale development and it has been extended well beyond the original scope of psychometric concerns (Duncan, 1984). To solidify construct validity, it is crucial to comprehensively evaluate the three fundamental types of validity: content validity, criterion-related validity, and construct validity (DeVellis, 2003). However, our review reveals a concerning trend: a majority of studies tend to emphasize content and construct validity, while criterion-related validity frequently does not receive the attention it deserves. This finding is consistent with previous research, where Kock et al. (2019) argue that the actual employment of newly developed scales to establish nomological validity is notably scarce in tourism literature. Therefore, there is an urgent need for future research to adopt a more balanced validation approach in the scale development process.

Additionally, the field of hospitality and tourism management research is a key area of behavioral science that often encounters the problem of method bias. The need to manage these biases during scale development, particularly during the criterion and nomological validation phases, cannot be overemphasized. To mitigate such bias, researchers are advised to implement both procedural and statistical remedies (Podsakoff et al., 2003). For instance, survey design should aim to eliminate item ambiguity and confusion. While collecting data for CFA or to analyze nomological validity, survey design strategies such as time lags, or changing the response prompts or presentation order of the constructs, can be employed to mitigate risks associated with response bias (Kao et al., 2020). Besides procedural control, statistical techniques such as Harman's single-factor test are also recommended to reduce the potential influences on the research findings. Method bias checks also help to determine whether a single-factor model or a multi-factor model is more suitable (Kao et al., 2020).

Furthermore, enhancing transparency is equally important, especially in reporting the initial number of items and the statistical software used in analyses. Smart-PLS, which is proficient in examining both reflective and formative measurement models, has been used in only 4.4% of studies. This aligns with the fact that most scales predominantly employ the reflective model. Such transparency is instrumental in facilitating the replication and verification of the scale’s findings.

Lastly, researchers are advised to meet or exceed the recommended minimum sample sizes to enhance the statistical robustness of their studies and consider employing multiple approaches to determine the necessary sample size for more complex scales.

Managerial and Theoretical Implications

It is our belief that the insights gained from these examinations and resultant suggestions can benefit researchers and practitioners in assessing various aspects of tourism-related phenomena at the global and local levels. Future survey research which examines tourism would enhance the validity and precision of its data by using scales which were developed in line with prominent scale development protocols (e.g. Churchill, 1979; DeVellis, 2003). The uniqueness and diversity of contexts in which tourism occurs suggests that future scale development is warranted to better attune the measurement of phenomena to the unique dynamics of these contexts. While such scales would sacrifice generalizability, they may also unveil boundary conditions and the influence of other factors which broadly oriented scales are not as well equipped to reveal.

The findings of the present study suggest that destination managers have a range of survey measurement tools at their disposal to gain insight from the perspective of tourists, local residents, and tourism organizations. Destination managers can assemble their own consumer research surveys by selecting scales which previous research suggests will likely be related. For instance, scales pertaining to tourist motivation can be used with those pertaining to tourist behavior. Destination managers could use scales together in this way to monitor performance and diagnose potential problems. For example, if visitation drops off for a tourism organization, the organization could survey its customers regarding their motivation. Then if their customers are lacking in a particular motivation which is important for the firm, such as novelty seeking or ego-enhancement, the issue could be addressed with targeted marketing messaging.

References

Churchill, J., and Gilbert, A. (1979). “A paradigm for developing better measures of marketing constructs.” Journal of Marketing Research, 16(1): 64-73.

Coltman, T., Devinney, T. M., Midgley, D. F., and Venaik, S. (2008). “Formative versus reflective measurement models: Two applications of formative measurement.” Journal of Business Research, 61(12): 1250-1262.

DeVellis, R. (2003). Scale Development: Theory and Applications, Applied Social Research Methods. Sage Publications. Thousand Oaks.

DeVellis, R. F., and Thorpe, C. T. (2021). Scale Development: Theory and Applications. Sage publications.

Duncan, O. D. (1984). Notes on Social Measurement: Historical and Critical. Russell Sage Foundation.

Edwards, J. R. (2011). “The fallacy of formative measurement.” Organizational Research Methods, 14(2): 370-388.

Gursoy, D., Uysal, M., Sirakaya-Turk, E., Ekinci, Y., and Baloglu, S. (2014). Handbook of Scales in Tourism and Hospitality Research. CABi.

Haley, S. M., and Osberg, J. S. (1989). “Kappa coefficient calculation using multiple ratings per subject: a special communication.” Physical Therapy, 69(11): 970-974.

Hinkin, T. R. (1995). “A review of scale development practices in the study of organizations.” Journal of Management, 21(5): 967-988.

Hinkin, T. R., Tracey, J. B., and Enz, C. A. (1997). “Scale construction: Developing reliable and valid measurement instruments.” Journal of Hospitality & Tourism Research, 21(1): 100-120.

Kao, C.-Y., Tsaur, S.-H., and Huang, C.-C. (2020). “The scale development of organizational culture on customer delight.” International Journal of Contemporary Hospitality Management, 32(10): 3067-3090.

Koc, E., and Ayyildiz, A. Y. (2022). “An overview of tourism and hospitality scales: Discussion and recommendations.” Journal of Hospitality and Tourism Insights, 5(5): 927-949.

Kock, F., Josiassen, A., and Assaf, A. G. (2019). “Scale development in tourism research: Advocating for a new paradigm.” Journal of Travel Research, 58(7): 1227-1229.

Landis, J. R., and Koch, G. G. (1977). “The measurement of observer agreement for categorical data.” Biometrics: 159-174.

Lin, H., Shi, S., and Gursoy, D. (2022). “Destination experiencescape: conceptualization and scale development amid COVID-19 pandemic.” Current Issues in Tourism, 25(24): 4047-4074.

Menor, L. J., and Roth, A. V. (2007). “New service development competence in retail banking: Construct development and measurement validation.” Journal of Operations Management, 25(4): 825-846.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, P. (2009). “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.” Annals of Internal Medicine, 151(4): 264-269.

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., and Ferreira, M. E. (2017). “Scale development: Ten main limitations and recommendations to improve future research practices.” Psicologia: Reflexão e Crítica, 30.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). “Common method biases in behavioral research: A critical review of the literature and recommended remedies.” Journal of Applied Psychology, 88(5): 879-903.

Schwab, D. (1980). “Construct validity in organizational behavior.” Research in Organizational Behavior, 2: 3-43.