Yu, S., & Hudders, L. (2022). Measurement invariance of the modified brand luxury index scale across gender, age and countries. Journal of Fashion Marketing and Management: An International Journal, 26(5), 870-889.
https://doi.org/10.1108/JFMM-10-2020-0235

Measurement invariance of the modified brand luxury index scale across gender, age and countries

Abstract

Many instruments have been developed to measure the perceived luxuriousness of brands, but one of the most frequently used scales is the 'brand luxury index' (BLI) from Vigneron and Johnson (2004) that distinguishes between high- and low-luxury brands. Despite its popularity and widespread use in academic research, the scale's psychometric properties and equivalence across cultures have been questioned. Recently, modified versions of the scale have been developed to strengthen the quality of the measurement. However, the performance and the measurement invariance of the modified version have not yet been investigated. The current paper aims to test the model fit of the modified BLI scale and the measurement invariance across gender, age, and country groups using nine datasets from a total of three different countries. The results of this analysis suggest that the modified BLI scale has an acceptable model fit and can be interpreted equivalently across gender and age groups. Metric invariance was found among the U.S., China, and India. However, scalar measurement invariance was established only across two countries: the U.S. and India. A follow-up analysis shows that partial scalar invariance can be established across the U.S., China, and India when removing constraints on the parameters of three items: Exclusive, Precious, Sophisticated.

Keywords: Brand luxury index, measurement invariance, perceived luxuriousness, measurement, scale

Introduction

Prior studies on the effectiveness of luxury marketing mainly investigate how consumers perceive luxury brands (e.g., Vigneron and Johnson, 2004), what motivation they have to purchase these brands (e.g., Han et al., 2010; Maziriri and Madinga, 2018), how luxury brands should be promoted to consumers (e.g., Parguel et al., 2016; Yu et al., 2017a; Yu et al., 2017b), and how willing consumers are to pay for luxury brands (Li et al., 2012). For example, many theories are employed to explain luxury consumption motives, such as the Veblen effect, bandwagon effect, social comparison theory, consumer culture theory, and self-concept theory (Ko, Costello, and Taylor, 2017).

Research on the definition and perception of luxury constitutes a foundation of luxury marketing research. Understanding the perception of luxury provides a key to measure the degree of perceived luxuriousness. For academia, an instrument to measure brand luxury can be useful when researchers would like to manipulate the brand type in the stimuli (e.g., luxury vs. non-luxury brand) (e.g., Yu, Hudders, and Cauberghe, 2018). Furthermore, brand luxury can be a very important dependent variable to assess the effectiveness of a marketing communication strategy. For example, Beuckels and Hudders (2016) investigated how image interactivity affected luxury perceptions of a brand in a virtual shopping environment, and Yu and Hu (2020) explored how showing Chinese (vs. Western) celebrities on Chinese social media influenced the perceived brand luxury.

Measuring the perceived luxuriousness of a brand is also beneficial for marketers in

luxury marketing. For example, based on the measured score, a brand can be possibly classified into different types: non-luxury brands (e.g., Zara), entry-level luxury brands (e.g., Hugo Boss), medium-level luxury brands (e.g., Moschino), top-level luxury brands (e.g., Louis Vuitton), and elite-level luxury brands (e.g., Puiforcat). In this way, a brand can be clearly positioned in the market.

Many scales have been developed to measure brand luxury (e.g., Hennigs, Wiedmann, and Klarmann, 2013). Among all of these instruments, one of the most-cited scales is the brand luxury index (BLI) developed by Vigneron and Johnson (2004). The BLI consists of five dimensions measuring the luxuriousness of the brand: *conspicuousness*, *uniqueness*, *quality*, *hedonism*, and *extended self*. The dimensions distinguish *high* or elite luxury brands from *low* or mass luxury brands.

Previous studies have tested the reliability of the BLI. Christodoulides, Michaelidou, and Li, (2009) showed some problems with the scale's psychometric properties and questioned its equivalence across cultures. For example, the reliability coefficients of conspicuousness and hedonism did not pass the threshold of 0.70. Exploratory and confirmatory factor analyses failed to provide support for the underlying structure of the BLI but indicated serious cross-loadings of items. The poor performance of the scale may be due to the design of the item. Respondents may find it difficult to rate a brand on a bipolar item ranging from unique to unusual as the nuance and difference between the two opposite ends of the scale may be unclear for respondents. Many researchers have revised the item to make it more distinct (e.g., Kim and Johnson, 2015; Doss and Robinson, 2013; Kluge et al., 2013; Parguel et al., 2016). A revised version of the BLI

scale has been found to offer better performance (Doss and Robinson, 2013). However, no tests of the reliability of the modified scale have been reported.

As the BLI scale is used more often, it is important to examine the reliability of the scale in different conditions. For example, does the measurement invariance of the BLI scale hold across different demographic groups such as gender, age and country? In other words, do people from different gender, age, and country groups understand and interpret the BLI scale in the same way. Previous studies (e.g., Stokburger-Sauer and Teichmann, 2013) have found that men and women have different understandings of uniqueness, status and hedonic values. Also, Hauck and Stanforth (2007) suggest that different age groups have different perceptions of luxury goods. A number studies (e.g., Kapferer and Michaut, 2016; Godey et al., 2013) also confirm that people from different countries use different terms to describe and define luxury. This implies that people in different groups may interpret the item in the BLI scale inequivalently. As such, measurement invariance of the BLI scale should be established to make meaningful comparisons across different demographic groups.

Nevertheless, studies on measurement invariance of the BLI scale across different demographic groups are, to our knowledge, non-existent. If people interpret the item differently across different groups, serious problems may occur when comparing the score across different groups. Research may lead to a biased conclusion if such issue is neglected. Therefore, it is necessary to test the measurement invariance of the BLI scale across different demographic groups.

Literature review

Measuring the perceptions of luxuriousness of the brand

The word *luxury* is everywhere. It appears on different types of products, from a bottle of lotion in a supermarket to a fancy hotel in Dubai. A bottle of mineral water can claim that it will give you a luxurious experience. Many generic brands claim to be luxury brands, while luxury brands rarely call themselves luxury. What does the phrase *luxury brand* mean? It is difficult to state a clear definition.

Luxury has been defined "as something more than necessary" (Mühlmann, 1975, p. 69). Dubois et al. (2001) defined luxury as something that is non-necessary and superfluous. Kapferer and Bastien (2012) described the function of the luxury product as to fulfill an individual's dream beyond needs or desires. Therefore, necessity is considered to be the key criterion to differentiate luxury goods.

In addition to *necessity*, *ordinariness* is another defining dimension. Luxury should be extraordinary; it is linked to expensiveness and exclusivity. Heine summarized all these definitions of luxury and defined luxury as "anything that is desirable and exceeds necessity and ordinariness" (Heine, 2012, p.46).

By definition, luxury brands are brands that produce luxury products. However, a luxury brand can also produce non-luxury products (e.g., Mercedes A-class cars). Therefore, Heine (2012) suggested that the categorization of a luxury vs. generic brand depends more on the brand image than the price.

The categorization of a brand is also determined by the country. For example, in China, Lacoste is considered a luxury brand, while it is regarded as a premium brand in France.

Fashion brands like Zara and Massimo Dutti try to blur the borderline between luxury and non-luxury brands by following a luxury strategy (e.g., opening stores next to luxury brands), which makes the difference between luxury and non-luxury brands even more confusing.

To position a brand within the continuum of luxuriousness, researchers have invented instruments to measure the luxuriousness of a brand. To develop such an instrument, the fundamental features of luxury brands need to be clearly identified. Chevalier and Mazzalovo (2015) identified three essential dimensions for luxury brands: artistry, craftsmanship, and internationalization. Table 1 summarizes the studies of many scholars on the perception of luxury (Dubois, Laurent and Czellar, 2001; Vigneron and Johnson, 2004; Kapferer and Bastien, 2012; De Barnier, Falcy and Valette Florence, 2012; De Barnier and Valette-Florence, 2013; Hennigs, Wiedmann, and Klarmann, 2013;).

Table 1. The literature on the perception of luxury brands

Dubois, Laurent and Czellar, 2001	Vigneron and Johnson, 2004	Kapferer and Bastien, 2012	De Barnier, Falcy and Valette Florence, 2012	De Barnier and Valette-Florence, 2013	Hennigs, Wiedmann, and Klarmann, 2013
Luxuriousness Superfluous & non-functional		Superlative		Functionality	Materialistic Usability
Make dream		Never comparative	Creativity	Aspiration	
Conspicuous Difference Elitist Refined people Reveal who you are	Conspicuousness (Elitism)		Elitism	Conspicuousness	Conspicuousness Prestige
High price			Power of the brand (preponderant	Expensiveness	Price

			position)		
Scarcity Uniqueness Few people own	Uniqueness	Rarity Exclusivity	Uniqueness		Uniqueness
Pleasure Pleasing	Hedonic (Refinement)		Hedonism Distinction Refinement	Self-pleasure	Hedonic
Not mass-produced Excellent quality	Quality	Hand-made Tradition	Quality	Premium Quality	Quality
Aesthetics Polysensuality Makes life beautiful		Fashion Arts Cultural mediation		Aesthetics	
Ancestral heritage & personal history		History Time	Renown	Personal History	
Good taste	Extended self (Power)				Self-identity
		Charity			

All scholars mentioned in Table 1 mentioned *quality*. Quality is commonly regarded as the most important feature of a luxury brand. In addition to quality, conspicuousness (4 times), non-functionality (4 times), uniqueness (or rarity, 5 times), expensiveness (3 times), aesthetics (3 times) and self-expression (3 times) are also considered to be important aspects of luxury brands. The dimensions described by Heine (2012) included price (i.e., expensiveness), quality (i.e., premium quality), aesthetics (i.e., a combination of beauty and elegance), rarity (i.e., limited productions), extraordinariness (i.e., a mind and style of its own) and symbolism (i.e., the best from the best). Ko, Costello, and Taylor (2017) proposed their own criteria of a luxury brand based on a review of the literature: 1. quality, 2. authentic value via desired benefits, 3. a prestigious image, 4. a premium price, 5. resonance with the consumer. By measuring these dimensions, one could possibly calculate the luxuriousness of a brand.

The brand luxury index scale

To measure perceived luxuriousness, Vigneron and Johnson (2004) developed the brand luxury index. The brand luxury index is one of the most popular scales used to measure perceived luxuriousness of a brand (Christodoulides, Michaelidou, and Li, 2009). The BLI scale contains 20 items that can be grouped into five dimensions: perceived conspicuousness, perceived uniqueness, perceived quality, perceived hedonism, and perceived extended self.

The first dimension, perceived conspicuousness, consists of 4 items: Conspicuous – Noticeable, Popular – Elitist (reverse-coded), Affordable – Extremely Expensive (reverse-coded), and For wealthy – For well-off. This dimension taps into the price perception and social status linked to the brand (Vigneron and Johnson, 2004).

The second dimension, perceived uniqueness, has also four items: Fairly exclusive – Very exclusive (reverse-coded), Precious – Valuable, Rare – Uncommon, and Unique – Unusual. The uniqueness dimension captures the perceptions of exclusivity and rarity, a driver of the desirability of a brand (Vigneron and Johnson, 2004).

The third dimension, perceived quality, includes five items: Crafted – Manufactured, Upmarket – Luxurious (reverse-coded), Best quality – Good quality, Sophisticated – Original, and Superior – Better. Perceived quality simply measures what factors lead consumers to perceive a brand as having the highest quality.

The fourth dimension is hedonism. This dimension has only three items: Exquisite – Tasteful, Attractive – Glamorous (reverse-coded), and Stunning – Memorable. The

hedonic dimension referred to the luxury perception "reflected by sensory gratification and sensory pleasure expected from the consumption" (Vigneron and Johnson, 2004, p. 491).

The last dimension is extended self. Four items form this dimension: Leading – Influential, Very powerful – Fairly powerful, Rewarding – Pleasing, and Successful – Well regarded. This dimension is related to the materialism and the construction of one's self in terms of luxury consumption and measures how a brand can express and represent one's personal success (Vigneron and Johnson, 2004).

This scale has been widely used by many researchers (e.g., Yu et al., 2018) and has been cited almost 2000 times as of 2021. Taking the volume of publications in luxury marketing into account, this scale has a great impact in luxury marketing. The popularity of this scale drives many scholars to re-evaluate its psychometric properties (e.g., Christodoulides, Michaelidou, and Li, 2009).

As the development of the BLI scale used a student sample in Australia, Christodoulides, Michaelidou and Li (2009) validated the scale using a sample of 260 luxury consumers in Taiwan. However, the results from the confirmatory factor analysis revealed an inadequate fit (Chi-square/df = 823/160, GFI = .76, RMSEA = .13, NFI = .79). The poor fit may be caused by different samples (students vs. luxury consumers), a different interpretation of the scale caused by cultural differences (Australian culture vs. Chinese culture), confusion regarding the bipolar answers (e.g., Uncommon vs. Rare), the translation (English vs. Chinese) or the response style. However, we cannot draw a conclusion on the reason(s) on the basis of their study. Conejo, Cunningham, and Young

(2019) conduct several studies to evaluate the performance of the BLI scale in US and China. Using both student and consumer samples, this study suggests that the BLI scale is factorially unstable. Also, only 30% of the items perform well. Although this study confirms the increasing concerns over this scale, it does not provide a solution to solve this issue.

The poor performance of the original BLI scale may be caused by the design of the item (Kim and Johnson, 2015). In particular, respondents sometimes find the BLI scale confusing because of the closeness of the bipolar answers. For example, a respondent from the U.S. once reported that she could not distinguish between the answers "unique" and "unusual". The closeness of the bipolar answers also makes translation of this scale extremely difficult. Therefore, many scholars have adapted this scale by making the two bipolar answers dissimilar (e.g., Kim and Johnson, 2015; Doss and Robinson, 2013; Kluge et al., 2013; Parguel, Delécolle, and Valette-Florence, 2016).

In one such revision, Doss and Robinson (2013) revised many items of the BLI scale. They changed Noticeable to Inconspicuous, Uncommon to Widely available, Upmarket to Practical, Better to Inferior, Original to Tacky, etc. Their revised BLI scale offers a better model fit than the original scale in the study of Christodoulides et al. (2009), but the fit values are still inadequate for luxury brands (GFI = .83, RMSEA = .98, NFI = .83) and acceptable for luxury counterfeit brands (GFI = .91, RMSEA = .06, NFI = .91).

Similar to Doss and Robinson (2013), Kim and Johnson (2015) modified some items in the BLI scale. For example, they replaced Rare – Uncommon with Very rare – Fairly rare, Unique – Unusual with Very unique – Fairly unique, etc., which makes the scale

much clearer for respondents. They also replaced the dimension "conspicuousness" with a new dimension, called "tradition" (Timeless – High fashion, Heritage – Emerging). The modified 13-item model yielded a good fit (GFI = .94, RMSEA = .05, NFI = .95).

Although many studies have attempted to re-evaluate the BLI scale, none have examined its measurement invariance. The absence of such an investigation may lead to inappropriate use of the scale. For example, measurement non-invariance may make mean scores of the brand luxury impossible to compare across gender, age or country groups. Misinterpretation of the results can lead to misleading conclusions.

Measurement invariance

As measurement is regarded as "*the systematic assignment of numbers on variables to represent characteristics of persons, objects, or events*" (Vandenberg and Lance, 2000, p. 4), measurement invariance, also known as measurement equivalence, is established to ensure cross-group comparisons (e.g., gender, ethnicity, culture, age) of mean differences, based on the measurement, are meaningful (Schmitt and Kuljanin, 2008). Measurement invariance refers to "'*whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute*" (Horn and McArdle, 1992, p. 117).

Vandenberg and Lance (2000) provided the first comprehensive review of papers on measurement invariance. This review demonstrates the importance of evaluation of measurement equivalence across populations. In it, they explained the definition and mechanism of different levels of measurement invariance and proposed a solution for

partial invariance. Furthermore, they proposed a comprehensive paradigm to help researchers to conduct such analyses in different contexts.

Schmitt and Kuljanin (2008) conducted an integrated review of 88 papers on measurement invariance published from 2000 to 2008. More and more researchers realized the importance of establishing measurement equivalence before testing the group difference. Measurement invariance can be applied to support the use of such instrument across age, gender, time, cohort, cultural/linguistic/racial groups (Schmitt and Kuljanin, 2008). Psychometric scales tested in these studies covered different areas, including psychology (e.g., depression, Motl et al., 2005), behavioral and organizational sciences (e.g., job satisfaction, Liu et al., 2014), education (e.g., academic motivation, Grouzet et al., 2006), individual differences (e.g., ethnocentrism, Yoo, 2002), and marketing (e.g., brand equity, Yoo and Donthu, 2001). In their study of marketing, Yoo and Donthu (2001) conducted a multi-group CFA to assess the measurement invariance of the brand equity scale in the U.S. and South Korea.

Steps in invariance testing

Vandenberg and Lance (2000) drew a flowchart of recommended steps for testing measurement invariance. In it, they recommended to first conduct an omnibus test of equality of covariance matrices. If the null hypothesis was not rejected ($\Sigma^{g} = \Sigma^{g'}$), it can be concluded that the overall measurement is invariant and further tests are not needed. However, if the null hypothesis was rejected, some levels of measurement invariance may still exist between groups (Vandenberg and Self, 1993).

First of all, it is necessary to conduct the configural invariance test; while the next step

is a test for metric equivalence (Horn and McArdle, 1992). If metric invariance is established, researchers can move on to the next step, a test of scalar invariance (Steenkamp and Baumgartner, 1998). If full metric invariance is not confirmed, the constraints can be released to check if partial measurement invariance exists.

However, according to Vandenberg and Lance (2000), tests for partial invariance can be quite exploratory, iterative and data-driven, which may pose threats to the integrity of results compared to a strict stepwise approach. Therefore, they suggested that the constraints can be relaxed under certain circumstances: 1. Only for a limited number of indicators; 2. Only built on a strong theoretical basis; 3. Supported by sufficient cross-validation evidence (Vandenberg and Lance, 2000, p. 38).

The proportion of scalar invariance among other measurement invariance tests is rather small; however, scalar invariance is an important prerequisite for comparing the latent mean of different groups (Vandenberg and Lance, 2000). Similar to the previous step, if scalar invariance cannot be established, researchers may check if partial scalar invariance holds across different groups. Next, with the establishment of scalar invariance, researchers may proceed to the next step involving a test of the invariance of the unique variances (uniquenesses) across different groups. If researchers are interested in structural invariance, they may conduct tests of invariances of factor variances, factor covariances and factor means. However, structural invariance is not the focus of this study; therefore, more attention will be paid to measurement invariance (the first five steps).

Level of measurement invariance

Configural invariance. Configural invariance is a test to see if the factors and pattern of factor loadings imposed on the items can explain the variance–covariance matrices (Schmitt and Kuljanin, 2008) and if such a factor structure is equivalent across different groups (Horn and McArdle, 1992). Failure to reject the null hypothesis means two groups follow the same conceptual frame of reference (Vandenberg and Lance, 2000). This means the factor structure implied is appropriate for both groups (*g* and *g*') and is identical between *g* and *g*'. Therefore, further tests of measurement invariance can proceed. In addition, the configural model is considered as a weak invariance test (Horn and McArdle, 1992) and serves as a baseline model for further analyses.

Metric invariance. Metric invariance is a test that constrains the factor loadings ($\Lambda_k^g = \Lambda_k^{g'}$). The *k*th indicator in the *g*th group of the CFA model can be represented by the following regression equation:

$$X_k^g = \tau_k^g + \Lambda_k^g \xi_k^g + \delta_k^g$$

In this equation, X_k^g is the *k*th observed item in the *g*th group; Λ_k^g refers to the factor loadings relating the X_k^g to the factor ξ_k^g; τ_k^g is the regression intercepts linking the observed measures to the underlying construct, that is, the observed value when ξ_k^g is set to zero; and δ_k^g represents the unique factor (uniquenesses).

The test of measurement invariance constrains the factor loadings across groups to be equal. The establishment of metric invariance is considered to be strong invariance according to Horn and McArdle (1992). A demonstration of metric invariance for items means researchers can compare the different values of those items across different groups and also that they can move on to the next test, for scalar invariance.

Scalar invariance. Many studies claim that measurement invariance has been established after achieving metric invariance (Vandenberg and Lance, 2000). However, if a study involves a comparison of latent means, demonstration of scalar invariance is required (Meredith, 1993). Scalar invariance is also sometimes called intercept invariance; the intercept is the value of the observed variable when the value of the latent construct is zero. The test of scalar invariance constrains the intercepts to be equal across all groups ($\tau_k^g = \tau_k^{g'}$). Establishing scalar invariance suggests responses that have the same value on the latent construct obtain the same score on the observed item across all groups (Milfont and Fischer, 2010).

Invariance of unique variances. Invariance of unique variances is also called residual invariance or error variance invariance. The invariance of unique variances involves a test to check if the sum of specific variance[1] and error variance is equal across different groups ($\delta_k^g = \delta_k^{g'}$).

Structural invariance. Previous studies have focused more on measurement equivalence than structural equivalence (Schmitt and Kuljanin, 2008). Structural invariance includes invariance of factor variances, factor covariances and factor means. Similar with measurement invariance, tests for structural invariance constrain the factor variances, factor covariances or factor means to be the same across different groups.

Measurement invariance across gender, age, and country groups

Ideally, a measurement instrument should be understood and interpreted in the same way. However, in reality, many factors may influence the measurement invariance. Differences in gender, age, and country may preclude responding to instruments in the

same way, and the items in the BLI scale may contain different meanings for different demographic groups. For example, the perceived uniqueness, status, and hedonic value are more prominent for women than men (Stokburger-Sauer and Teichmann, 2013). In other words, the items like “unique”, “conspicuous”, and “rewarding” may have stronger weight on the latent construct for women than men.

Furthermore, Hauck and Stanforth (2007) suggested that different cohort groups have different perceptions of luxury goods. The older cohorts tend to view more categories of goods (e.g., cell phone) to be a luxury than the younger cohorts. Therefore, age can also be a potential factor influencing the measurement invariance.

In addition to gender and age, the country of the respondent may also strongly influence the comparability of the scale because people from different countries may have different perception of luxury (e.g., Kapferer and Michaut, 2016; Godey et al., 2013). Kapferer and Michaut (2016) conducted a cross-cultural study on consumers’ perception of luxury and investigated which attributes are used in each country to define luxury. The respondents in each country were asked to select four characteristics that best described “luxury” in their view from 14 potential defining attributes of luxury listed by researchers. The frequency with which any of the 14 proposed items are selected indicates its importance perceived by the respondents as to define luxury. According to the results, “high quality,” “expensiveness,” and “prestige” are the top three items selected to define luxury, composing the core attributes of luxury worldwide. Nevertheless, people from different countries give different weight to attributes that best define luxury, their perceptions of luxury vary on a country level (Table 2). For

example, people in Japan regard quality as the first defining attribute of luxury, while people in China are the only group taking expensiveness as the prime defining attribute, and people in Brazil pay more attention to pleasure when evaluating luxury. This implies that the latent construct of brand luxury may have different factor loadings on an item (e.g., quality) in different countries. Therefore, measurement invariance of the BLI scale should be carefully examined to avoid misleading interpretation of the comparison across different demographic groups.

Table 2: Attributes selected to best define luxury around the world (Kapferer and Michaut, 2016)

	France	USA	China	Brazil	Germany	Japan	Total
High quality	60%	69%	45%	61%	57%	75%	60%
Expensive	41%	53%	57%	31%	46%	40%	45%
Prestige	55%	48%	40%	31%	32%	56%	43%
Pleasure	38%	40%	15%	42%	34%	14%	30%
Beauty	30%	25%	12%	33%	30%	32%	26%
Fashion	21%	25%	44%	22%	30%	12%	27%
Dream	36%	20%	14%	32%	36%	11%	25%
Minority	17%	16%	36%	33%	24%	18%	25%
Rarity	29%	17%	26%	22%	18%	32%	24%
Personalized service	16%	19%	31%	33%	20%	9%	22%
Heritage	15%	12%	24%	14%	13%	41%	20%
Timeless	14%	23%	19%	5%	25%	29%	19%
Art	17%	14%	22%	19%	19%	25%	19%
Innovation	10%	19%	15%	22%	16%	7%	14%

Methods

Data

Nine datasets were used in this study (Table 3). The data were collected across three countries, the United States (5 datasets), China (2 datasets), and India (2 datasets) from

2016 to 2018. A typical dataset includes 20 questions to measure perceived luxuriousness, 5 questions to measure consumers' attitude toward the brand, 10-15 questions about the participants (e.g., materialism, need for uniqueness), and 5-7 questions about the demographic information. We have selected these nine datasets for the following reason. First, the data were collected in U.S., China, and India. The U.S. and China are the two largest markets for luxury goods in the world and India is one of the most important emerging markets for luxury products (Deloitte, 2018). These three countries differ greatly in culture, languages, economic development, and political systems. We sought to determine if people from such distinctive countries interpret the BLI scale in a conceptually similar manner.

The nine datasets were originally used for academic purposes and some of the datasets have been published in international peer-reviewed journals. In particular, respondents were recruited in the same period and their number was roughly equal in both genders. Furthermore, respondents were recruited through online panels including Mechanical Turk, Qualtrics, and the Baidu Survey Centre. A Qualtrics panel is clearly superior in research demanding for representativeness or sample diversity on demographic variables (Boas, 2020). Mechanical Turk has become widely used by researchers in social sciences and it has been examined that respondents on Mechanical Turk provide valid measures of variables that are important to social scientists (Johnson and Ryan, 2020). For Baidu Survey Centre, it is a leading panel service provider in China. With over 17 million registered users, it provides researchers with a wide-ranged online panel (Baidu, 2021).

Third, the datasets contained different types of brands. Respondents were asked to rate the perceived luxuriousness of a brand. These brands were all well-known luxury brands or fashion brands except datasets 5, 7, and 9, which were collected using a fictitious luxury brand, “Dulcet”. Dior, Louis Vuitton and Burberry are popular luxury brands around the world. They rank the fourth, fifth and fourteenth respectively in the “The 15 Most Popular Luxury Brands Online in 2021” by Luxe Digital (2021), as each enjoys significant customer attention and drives substantial online conversations. Zara and H&M both have a vast market and consumer base across nations. Zara has more nearly 3,000 stores in 96 countries as of 2019, China (229 stores), the U.S. (98 stores), India (22 stores) and has generated 21.9 billion dollars in 2019’s annual sales (Forbes, 2020). H&M has 5,076 stores worldwide as of 2019, the U.S. (593 stores), China (520 stores) and India (48 stores). H&M has 114 million members in its customer loyalty program and has achieved 187 billion Swedish kronor in net sales in 2020 (H&M Group, 2021).

The language of the questionnaire was English for American and Indian respondents and Chinese for Chinese respondents. We did not translate the English questionnaire into Hindi because English as one of the two national languages of India is widely used in India. According to Statista (2020), the number of English internet users in India was about 175 million in 2016 and was estimated to reach 199 million in 2021. In addition, a survey about the share of English speakers across India shows that around 88 percent respondents in urban areas spoke English (Keelery, 2020). The Chinese version was translated by a Chinese-speaking student and back-translated by another native speaker

with very high English proficiency, who has received a master's degree in English literature and has rich experience in translation work.

We merged all datasets (datasets 1–9) for use in testing measurement invariance across gender. For testing measurement invariance across different age groups, only datasets 2–9 were merged, because dataset 1 contained no age information. With regard to tests of measurement invariance across countries, we selected two datasets to represent each country. For each country, one dataset contained a real brand, while another dataset was collected using the same fictitious brand. Datasets 2 and 5 were used to represent the U.S. and dataset 6 and 7 for China. Dataset 8 and 9 were used for the Indian group. The sample size in each country was similar (Table 4). The sample size balanced across different gender, age, and countries.

Table 2. A description of the dataset used

NO. dataset	Country	Brand	Sample size	Mean age	Male ratio	Language
1	U.S.	Louis Vuitton /Zara	125	n/a	47.2%	English
2	U.S.	Burberry /H&M	182	35.8 (11.5)	51.6%	English
3	U.S.	Burberry	278	39.4 (12.6)	44.2%	English
4	U.S.	Dior	58	37.2 (13.1)	51.7%	English
5	U.S.	Dulcet (fictitious)	105	33.8(10.6)	64.8%	English
6	China	Louis Vuitton	118	25.4 (4.1)	29.7%	Chinese
7	China	Dulcet (fictitious)	160	29.7 (6.5)	46.3%	Chinese
8	India	Burberry /H&M	110	29.3 (6.0)	74.5%	English
9	India	Dulcet (fictitious)	72	27.7 (3.8)	75.0%	English

Table 3. The sample size for each test of measurement invariance and demographic information

Test		Sample size	Percentage
gender	Male	617	0.51
	Female	591	0.49
	Sum	1208	1
age	18-25	251	0.23
	26-35	510	0.47
	36-55	250	0.23
	56 or older	72	0.07
	Sum	1083	1
country	U.S.	287	0.38
	China	278	0.37
	India	182	0.25
	Sum	747	1
Monthly household income (net) in dollar	$0 - $1,000	260	0.22
	$1,001 - $2,000	197	0.16
	$2,001 - $3,000	190	0.16
	$3,001 - $4,000	102	0.09
	$4,001 - $5,000	84	0.07
	$5,001 - $6,000	88	0.07
	$6,001 - $7,000	73	0.06
	$7,001 - $8,000	40	0.03
	$8,001 - $9,000	23	0.02
	> $9,000	146	0.12
	Missing values	5	0.00
	Sum	1208	1
Education level	Less than High School	17	0.01
	High School	234	0.19
	Bachelor's Degree	762	0.63
	Master's Degree	178	0.15
	Doctoral Degree	17	0.01
	Sum	1208	1

Measures

The brand luxury index contains 20 items (Table 5). As mentioned earlier, we used a modified version of the brand luxury index adapted from the studies of Doss and

Robinson (2013) and Kim and Johnson (2015). These two modified versions are quite similar; confusing items (e.g., Conspicuous) were replaced by terms that were more clearly bipolar.

However, these two modified versions are also not perfect. For the modified 20-item BLI scale proposed by Kim and Johnson (2015), the difference between the two bipolar items remains too small. For example, it is difficult to tell the difference between exquisite and tasteful, superior and better. For the other scale, Doss and Robinson (2013) suggested that when using their modified BLI scale, fit values were inadequate for luxury brands. Therefore, in this study, a combination of the two modified versions was used to avoid these issues. The whole selection procedure was listed as follows:

1. The revised 20 items from the study by Kim and Johnson (2013) were adopted.

2. We identified both the well-modified items (2, 3, 9), items that need to be further modified (1, 5, 7, 8, 15, 16, 18), and problematic items (4, 6, 10, 11, 12, 13, 14, 17, 20).

3. Well-modified items were adopted without further revision. For items that need further modification, we replaced the term “fairly” with “not” to make the difference between the two bipolar items more prominent.

4. For problematic items, we used the items (4, 6, 11, 12, 13, 17, 20) modified by Doss and Robinson (2013). For item 10 and item 14, we further adapted the item. For item 10, Doss and Robinson (2013) used the adjective “practical” as the opposite of “luxurious”. We believe the term “practical” includes many other meanings, such the functional value. Therefore, we simply used “not luxurious”. For item 14, Doss and Robinson (2013) changed the adjective “exquisite” to “beautiful” and “tasteful” to

"ugly", which deviates from the original meaning. Therefore, we kept the item "exquisite" and added "not exquisite" as the opposite item.

Table 4. The revised items in the current study

Original items	Revised items
Luxe_1 Conspicuous – noticeable [r]	Not noticeable – very noticeable*
Luxe_2 Popular – elitist	Popular to public – elitist*
Luxe_3 Affordable – extremely expensive	Affordable – extremely expensive*^
Luxe_4 For wealthy – for well-off [r]	For the non-wealthy – For wealthy^
Luxe_5 Fairly exclusive – very exclusive	Not exclusive – very exclusive*
Luxe_6 Precious – valuable [r]	Disposable – very precious^
Luxe_7 Rare – uncommon [r]	Not rare – very rare*
Luxe_8 Unique – unusual [r]	Not unique – very unique*
Luxe_9 Crafted – manufactured [r]	(Mass) Manufactured – (hand) crafted*^
Luxe_10 Upmarket – luxurious	Not luxurious – very luxurious
Luxe_11 Best quality – good quality [r]	Poor quality – best quality^
Luxe_12 Sophisticated – original [r]	Tacky – very sophisticated*
Luxe_13 Superior – better [r]	Inferior – very superior^
Luxe_14 Exquisite – tasteful [r]	Not exquisite – very exquisite
Luxe_15 Attractive – glamorous	Not glamorous – glamorous^
Luxe_16 Stunning – memorable [r]	Not stunning – stunning*
Luxe_17 Leading – influential [r]	Not leading – very leading^
Luxe_18 Very powerful – fairly powerful [r]	Not powerful – very powerful*^
Luxe_19 Rewarding – pleasing [r]	Not rewarding – very rewarding^
Luxe_20 Successful – well-regarded [r]	Unsuccessful – very successful^

Note: *modified by Kim and Johnson (2015); ^ modified by Doss and Robinson (2013)

Statistical Analysis Procedures

Model estimation and procedure

Analyses were conducted using AMOS version 23 using the maximum likelihood estimation procedure. As the questionnaire was conducted online using Qualtrics, the

force response function provided a solution for avoiding missing values.

First, following the flowchart proposed by Vandenberg and Lance (2000), an omnibus test of equivalence of the groups' covariance matrices was conducted before starting a multi-group CFA (Finch and West, 1997). The structures in the tested groups did not differ significantly if the χ^2 was not significant; further measurement invariance analysis will not be necessary because measurement invariance can be established if covariance matrices do not differ across groups (e.g., Bagozzi & Edwards, 1998; Cole & Maxwell, 1985; Steenkamp & Baumgartner, 1998). If χ^2 was significant, we need to locate the level of measurement non-invariance. In this case, we moved on to the next step, which was the multi-group CFA. Next, we conducted tests of configural, metric, scalar, and residual invariance one by one because invariance in the prior step must be established as a necessary condition for the evaluation of further aspects of measurement invariance of the next step (Vandenberg and Lance, 2000). If a lower level of measurement invariance was established, a test for a higher level can be continued. For example, if configural invariance is not established, a test of metric invariance cannot be warranted because this implies that the observed measures have different constructs in each group (Vandenberg and Lance, 2000). Following the suggestion of Vandenberg and Lance (2000), if measurement invariance at a certain level cannot be established, we examined whether partial invariance exists. If not, we stopped without conducting further tests and reported the level of measurement invariance.

Criteria for the evaluation of measurement invariance

Overall model fit refers to "evaluating the ability of the *a priori* model to reproduce the

observed matrix" (Vandenberg and Lance, 2000, p.43). The evaluation of configural invariance relies on the indexes of the goodness of fit. A good fit of the configural model suggests that the measures are anchored to the same configuration for each group (Vandenberg and Lance, 2000). One of the most frequently used criteria is the chi-square test of model fit. An insignificant chi-square value provides support for a well-fitting model. However, many studies have found that the chi-square value is very sensitive to the sample size. A large sample size can lead to higher chi-square value (Hu and Bentler, 1993; Bollen, 1989). Instead of the chi-square, as suggested by Kline (2015), this study will use SRMR, RMSEA, and CFI to evaluate the goodness of fit of the model. The cut-off points for each model fit index are as follows:

Table 5. Cut-off points proposed by previous studies

Model fit index	Value range	Critical values	Reference
SRMR	[0;1]	SRMR ≤.08: good fit	Kline, 2015
RMSEA	[0;∞]	RMSEA < .05: good fit RMSEA > .08: mediocre fit RMSEA > .10: poor fit	Byrne, 2016
CFI	[0;1]	CFI ≥ .90: good fit	Kline, 2015

For comparing the different model, we used the recommendations offered by Chen (2007) to determine cut-off points for metric invariance. In Chen's method, when the sample size is greater than 300, the assumption of metric (scalar and residual) invariance is rejected when a change in CFI is greater than -.01 (-.01 for scalar and residual invariance), supplemented by changes in RMSEA greater than -.015 (-.015 for scalar and residual invariance) or changes in SRMR greater than -.03 (-.01 for scalar and residual).

Results

Overall model specification

We first conducted a CFA with all data to examine the model fit. The results indicated an acceptable fit of the model (CMIN/df = 1504.09/160 = 9.40; CFI = .922; TLI = .907; SRMR = .056; RMSEA = .083). However, all factor loadings were positive and substantial (higher than .60) except the item, *Noticeable* (Luxe_1), having a standardized factor loading of .28. This item has been reported as problematic in prior studies (e.g., Doss and Robinson, 2013). After deleting this item, the overall model fit improved (CMIN/df = 1317.25/142 = 9.28; CFI = .931; TLI = 0.917; SRMR = .043; RMSEA = .083). Items like Expensive, For wealthy, Exclusive, Precious, Luxurious, Superior, Exquisite, Stunning, Powerful, Leading had relatively high factor loadings (Above .70). In the following analysis, we used the 19-item model.

Measurement invariance across gender groups

An omnibus test of equivalence of the gender groups' covariance matrices obtained a significant χ^2 (290.9/190, $p < .001$), indicating that the structure differs in both groups. Accordingly, we conducted a multi-group CFA. The first test was of the configural invariance, a pre-condition for metric invariance. The configural or unrestricted model (Model 0) had a good fit (CMIN/df = 1524.2/284, RMSEA = .060, CFI = .927, SRMR = .043), which suggested that groups of male and female respondents had the same structure. Next, Model 1 constrained all the factor loadings to be equal in each group. Compared with Model 0, the model fit of Model 1 (metric invariance) was not significantly worse (ΔCFI = 0, ΔRMSEA = .001, ΔSRMR =.001) according to Chen

(2007). Beside the factor loadings, Model 2 restricted the intercepts to be equal in each group. The results suggested that the measurement satisfied the scalar invariance test (ΔCFI = .002, ΔRMSEA = .001, ΔSRMR =.000). As a result, the means could be compared across genders. Furthermore, the test of residuals, factor variances, and factor covariances invariance suggested that the uniquenesses, factor variances, and factor covariances were equivalent across two gender groups (Table 7).

Table 6. Model fit of measurement invariance across gender

Model	CMIN	DF	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
0.Configual	1524.2	284	.927	.060	.042			
1.Loadings	1538.9	298	.927	.059	.043	.000	.001	.001
2.Intercepts	1599.0	317	.925	.058	.043	.002	.001	.000
3.Residuals	1644.8	336	.923	.057	.045	.002	.001	.002
4. Factor variance	1648.6	346	.924	.056	.046	.001	.001	.001
5. Factor covariance	1651.6	351	.924	.055	.046	.000	.001	.000

Measurement invariance across age groups

The general test of equality of the age groups' covariance matrices shows that the χ^2 (1056.6/570, $p < .001$) was highly significant. A further analysis of the configural model (Table 8) shows that it produced a good fit index for all groups (CMIN/df = 1877.3/568, CFI = .916, RMSEA = .046, SRMR =.046). Next, we held the items' weight equal across all the age groups. Metric invariance was achieved (ΔCFI = .001, ΔRMSEA = .001, ΔSRMR =.002). After the metric invariance was obtained, we constrained the intercepts and residuals of all items to be equal across groups. Tests for the age group invariance of the item intercepts and residuals demonstrated that the changes in CFI,

RMSEA, SRMR were still acceptable (Intercepts: ΔCFI = .009, ΔRMSEA = .000, ΔSRMR =.000; Residuals: ΔCFI = .008, ΔRMSEA = .000, ΔSRMR =.007). Therefore, the measurement invariance was established. However, the factor variances' invariance did not hold across all the groups because the change of the SRMR was greater than the cut-off point (ΔCFI = .001, ΔRMSEA = .000, ΔSRMR =.026).

Table 7. Model fit of measurement invariance across four age groups

Model	CMIN	DF	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
0.Configual	1877.3	568	.916	.046	.046			
1.Loadings	1925.0	610	.915	.045	.048	.001	.001	.002
2.Intercepts	2126.0	667	.906	.045	.048	.009	.000	.000
3.Residuals	2300.3	724	.898	.045	.055	.008	.000	.007
4. Factor variance	2339.5	739	.897	.045	.081	.001	.000	.026
5. Factor covariance	2425.4	769	.893	.045	.076	.004	.000	.005

Measurement invariance across three country groups

To conclude, we tested the measurement invariance of a modified BLI across three countries, the U.S., China, and India. The omnibus test of equality of the covariance matrices produced a significant chi-square, suggesting the covariance matrices were not equivalent (χ^2 = 514.9/190). Consequently, a multi-group CFA was conducted. The configural invariance model provided a good model fit (CFI = .916, RMSEA = .054, SRMR = .035).

Next, we compared the metric invariance model with the configural model. The changes in CFI, RMSEA and SRMR were smaller than the cut-off value (ΔCFI = .001, ΔRMSEA = .001, ΔSRMR =.002). The next model tested whether the item intercepts were invariant across countries. Model fit results showed that ΔCFI was larger than the

cut-off value (ΔCFI > .01). Therefore, scalar invariance did not hold across three countries. In other words, the latent means could not be compared across the U.S., China, and India.

A possible solution was to check if partial measurement invariance existed. According to Vandenberg and Lance (2000), partial measurement equivalence is acceptable only when the number of parameters relaxed to vary across groups is relatively small. Furthermore, partial measurement invariance should be built on a strong theoretical foundation.

In our case, one possible reason for intercept non-invariances may be translational non-equivalence. In the U.S. and India, the scale was in English, while in China, the scale was translated into Chinese. During the translation procedure, the two translators felt that it was difficult to translate certain items into Chinese because the corresponding word did not exist in Chinese or it may have different meanings. For example, the term Exclusive was translated as "独有的", while Disposable and Sophisticated were translated as "可随意处置的" and "高雅的", respectively. These Chinese terms cannot fully convey the original meaning in English and may imply other meanings which might influence the interpretation. Therefore, the model of partial measurement invariance may release the restriction of the parameters on these items (i.e., Luxe_5 Exclusive, Luxe_6 Precious, Luxe_12 Sophisticated) for China. After allowing the parameters of these items' intercepts to vary in China while still constraining the parameter across U.S. and India (i5_US = i5_India, i6_US = i6_India, i12_US = i12_India), the model fit improved compared to Model 1. The ΔCFI was smaller than

the cut-off point (ΔCFI = .009 < .01) proposed by Chen (2007).

Table 8. Model fit of measurement invariance across the U.S., China, and India

Model	CMIN	DF	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
0.Configual	1363.0	426	.916	.054	.035			
1.Loadings	1400.0	454	.915	.053	.037	.001	.001	.002
2.Intercepts	1611.0	492	.899	.055	.040	.016	.002	.003
2a. Partial	1513.82	489	.906	.054	.042	.009	.001	.001
3.Residuals	1842.8	530	.882	.058	.046	.024	.004	.004
4. Factor variance	1964.2	540	.872	.060	.099	.010	.002	.053
5. Factor covariance	2002.2	560	.870	.059	.103	.002	.001	.004

To sum up, weak measurement invariance (metric) was supported across the U.S., China, and India. Strong measurement invariance could be established across the U.S. and India. However, strong (scalar) measurement invariance did not hold across all three countries. Only partial scalar invariance that allowed the parameters of three items (Exclusive, Precious, Sophisticated) to vary in China could be established.

Discussion and implications

The year 2020 was marked by the Covid-19 pandemic, which had a significant impact on luxury industry. The market for personal luxury goods contracted for the first time since 2009, dropping by 23% at current exchange rates to hit €217 billion, and the overall luxury market decreased by 20% to 22% at current exchange rates, to an evaluated €1 trillion (Bain & Company, 2021). Covid-19 has profoundly changed the way people live, the way they shop and value things. When people are under quarantine, how, when and why they buy luxury products have changed accordingly. After the

luxury market has taken a hit by the pandemic, many luxury brands (e.g., Chanel, Louis Vuitton) increase the price of their product to strengthen their luxury perception (McInnis, 2021). For luxury brands, maintaining a high level of luxuriousness is very important, especially during the crisis. The present study provides a solution for both researchers and practitioners to measure the perceived luxuriousness of the brand.

This paper examined the model fit of the modified BLI scale and its measurement invariance across gender, age and countries. This study first shows that the modified 19-item BLI scale has a good model fit. Tests of the invariance of the brand luxury index scale further suggested that the BLI scale can be interpreted equivalently across gender and age groups. Weak (metric) measurement invariance was found among three countries; strong (scalar) measurement invariance holds only across the U.S. and India, not for China. Only after releasing the constraints on three items could partial scalar invariance could be established across three countries.

The scalar non-invariance could be caused by the translation of the instrument; translation from English to Chinese may distort the original meaning of the item. In this paper, the scales had two language versions: Chinese and English. Other studies have shown that translated measures are more likely to lack equivalence across different language versions (Schmitt and Kuljanin, 2008). A translation explanation is supported by the fact that strong measurement invariance was found across the U.S. and India, areas where an English version of the scale was used, and after removing the constraint on the parameter of three items that may have experienced translational problems, partial scalar invariance was established across all three countries.

This study is the first study to test the model fit of the modified BLI scale. The findings of this paper contribute to both the academia and industry. We recommend scholars and marketers use a modified 19-item BLI scale to measure the perceived luxuriousness of brands in future research. First, the modified BLI scale tested in the current study offers very good performance with model fit values of a quality that has rarely been seen in prior research. The original scale of Vigneron and Johnson (2004) has been criticized for its poor model fit (Christodoulides et al., 2009). The modified scale of Doss and Robinson (2013) also has problems with the fit value. Second, the modified 19-item scale also shows adequate measurement invariance across different gender, age, and countries. For scholars and marketers, the establishment of the metric invariance of the modified 19-item BLI scale implies that the scale can be used across gender, age and countries (U.S., China, and India) if the purpose of the study is to understand the relationship between some variables and perceived luxuriousness of a brand. However, for researchers that would like to compare the score of brand luxuriousness across different countries such as China, U.S. and India, we recommend them to use the modified 16-item BLI scale (without Noticeable, Exclusive, Precious, Sophisticated) because its scalar invariance is established. Furthermore, for marketers that would like to track consumers' luxury perception about a brand overtime, a 19-item scale may not be handy for them. In this case, marketers may select items with high factor loadings, such as Expensive, For wealthy, Exclusive, Precious, Luxurious, Superior, Exquisite, Stunning, Powerful, and Leading. This 10-item scale also consists of five dimensions and each dimension has two items. However, this short form of the BLI scale needs

further validity tests to see if it can perform as well as the 19-item one. Future studies may even work on a shorter version (e.g., 5 items) and test its validity.

Additionally, researchers may not need to worry about the measurement invariance of the modified BLI scale across gender and age groups. With a large sample size, this study has shown robust evidence of measurement invariance across gender and age groups. However, when using this scale across different countries, a monolingual questionnaire seems to be a better option to avoid translation-based non-equivalence. Translation can produce additional meanings or reduce the original meaning, which leads to measurement non-equivalence (Schmitt and Kuljanin, 2008). However, in reality, not all respondents can understand the same language. If it is necessary to use a translated version of BLI, researchers should carefully consider the choices related to translation and must conduct back-translation. In addition, researchers should also conduct tests of measurement invariance across different language versions to locate any problematic items.

Limitations and future research

Although this study is the first to examine the measurement invariance of the BLI scale across different demographic groups, it entails several limitations. For the tests of measurement invariance across the U.S., China, and India, the brand used in each country was different (China: Louis Vuitton; U.S./India: Burberry, H&M). Therefore, the scalar non-invariance could have resulted not just from language translation but also from the difference in brands, i.e. the influence of the country on measurement

invariance may be confounded with the difference of the brand. It will be necessary for future research to use the same brand when testing the measurement invariance to avoid this issue.

Moreover, tests of measurement invariance across countries can be improved by strictly controlling for other potentially confounding factors such as the translation, stimuli, sampling, and response styles. For example, in this study, we used the English version for respondents from U.S. and India. Future research may use a Hindi version to see if the same level of measurement invariance still holds.

Researchers may be interested in extending the findings of this study to more countries. Apart from the three countries examined in this study, Japan, South Korea, and the United Arab Emirates, each with a distinct culture, are also important markets for luxury goods.

Furthermore, in addition to gender, age and countries, other demographic groups such as income level and educational background may provide an interesting focus for further study. Also, this study did not set restrictions on the purchase history or knowledge of luxury brands of the respondents, it is advisable that further studies could recruit luxury consumers or even examine the measurement invariance between regular and irregular luxury consumers.

It should not be neglected that although measurement variance of the BLI scale can be established across many groups, using a standardized and simplified scale to fit all situations is impossible and risky. For example, an attempt to use the BLI scale for all countries is not appropriate for marketers and scholars to fully capture local consumers'

real perception of luxuriousness because consumers in different counties have different interpretation of luxury. A better solution is to create a scale for each country, such as a BLI-US, a BLI-Brazil, and a BLI-Japan.

Last but not least, the data for this study was collected before Covid-19. The pandemic may have changed people's luxury consumption behavior or even the opinion about luxury products. Future studies may conduct a longitudinal analysis to examine the measurement invariance across different periods.

Notes

1. Putnick and Bornstein describe specific variance as "variance of the item that is not shared with the factor" (2016, p. 77).

References

Baidu (2021). (2021). About Baidu MTC. Retrieved from https://mtc.baidu.com/.

Bain & Company. (2021). The Future of Luxury: Bouncing Back from Covid-19. Retrieved from https://www.bain.com/insights/the-future-of-luxury-bouncin-back-from-covid-19/.

Beuckels, E., & Hudders, L. (2016). An experimental study to investigate the impact of image interactivity on the perception of luxury in an online shopping context. *Journal of Retailing and Consumer Services*, *33*, 135–142.

Boas, T., Christenson, D., & Glick, D. (2020). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods, 8*(2), 232-250.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303–316.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chevalier, M., & Mazzalovo, G. (2015). *Management et Marketing du luxe-3e éd.* Dunod.

Christodoulides, G., Michaelidou, N., & Li, C. H. (2009). Measuring perceived brand luxury: An evaluation of the BLI scale. *Journal of Brand Management, 16*(5–6), 395–405.

Conejo, F. J., Cunningham, L. F., & Young, C. E. (2020). Revisiting the Brand Luxury Index: New empirical evidence and future directions. *Journal of Brand Management, 27*(1), 108–122. https://doi.org/10.1057/s41262-019-00168-4

De Barnier, V., & Valette-Florence, P. (2013). Culture and luxury: An analysis of luxury perceptions across frontiers. In *Luxury marketing* (pp. 37–56). Springer.

De Barnier, V., Falcy, S., & Valette-Florence, P. (2012). Do consumers perceive three levels of luxury? A comparison of accessible, intermediate and inaccessible luxury brands. *The Journal of Brand Management*. 19. 10.1057/bm.2012.11.

Deloitte. (2017). *CHINA LUXURY E-COMMERCE WHITEBOOK*.

Doss, F., & Robinson, T. (2013). Luxury perceptions: Luxury brand vs counterfeit for young US female consumers. *Journal of Fashion Marketing and Management: An International Journal*, *17*(4), 424–439. https://doi.org/10.1108/JFMM-03-2013-0028

Dubois, B., Laurent, G., & Czellar, S. (2001). *Consumer rapport to luxury: Analyzing complex and ambivalent attitudes*. HEC Paris.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality*, *31*(4), 439–485.

Forbes. (2020). Zara. Retrieved from https://www.forbes.com/companies/zara/?sh=3d25cc067487.

Godey, B., Pederzoli, D., Aiello, G., Donvito, R., Wiedmann, K., & Hennigs, N. (2013). A cross-cultural exploratory content analysis of the perception of luxury from six countries. *Journal of Product & Brand Management*, *22*(3), 229–237. https://doi.org/10.1108/JPBM-02-2013-0254

Grouzet, F. M., Otis, N., & Pelletier, L. G. (2006). Longitudinal cross-gender factorial invariance of the Academic Motivation Scale. *Structural Equation Modeling*, *13*(1), 73–98.

Han, Y. J., Nunes, J. C., & Drèze, X. (2010). Signaling Status with Luxury Goods: The Role of Brand Prominence. *Journal of Marketing*, *74*(4), 15–30. https://doi.org/10.1509/jmkg.74.4.15

Hauck, W. E., & Stanforth, N. (2007). Cohort perception of luxury goods and services. *Journal of Fashion Marketing and Management: An International Journal*, *11*(2), 175–188. https://doi.org/10.1108/13612020710751365

Heine, K. (2012). The concept of luxury brands. *Luxury Brand Management*, *1*, 2193–1208.

Hennigs, N., Wiedmann, K.-P., & Klarmann, C. (2013). Consumer value perception of luxury goods: A cross-cultural and cross-industry comparison. In *Luxury marketing* (pp. 77–99). Springer.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144.

Hu, L.-T., Bentler, P. M., & Hoyle, R. H. (1995). Structural equation modeling: Concepts, issues, and applications. *Evaluating Model Fit*, 76–99.

H&M Group. (2021). Annual report 2020.

Johnson, D., & Ryan, J. (2018). Amazon Mechanical Turk Workers Can Provide Consistent and Economically Meaningful Data. MPRA Paper 88450. University Library of Munich, Germany.

Kapferer, J.-N., & Bastien, V. (2012). *The luxury strategy: Break the rules of marketing to build luxury brands*. Kogan Page Publishers.

Kapferer, J.-N, & Michaut, A. (2016). Pursuing the Concept of Luxury A cross-country comparison and segmentation of luxury buyers' perception of luxury. *Journal of International Marketing Strategy*, 4. 6-23.

Keelery, S. (2020). Number of Indian and English language internet users in India 2011-2021. In *Statista*. Retrieved from https://www.statista.com/statistics/718420/internet-user-base-by-language-india/.

Kim, J., & Johnson, K. K. (2015). Brand luxury index: A reconsideration and revision. *Journal of Fashion Marketing and Management: An International Journal*, *19*(4), 430–444. https://doi.org/10.1108/JFMM-05-2015-0043

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kluge, P. N., Königsfeld, J. A., Fassnacht, M., & Mitschke, F. (2013). Luxury web atmospherics: An examination of homepage design. *International Journal of Retail & Distribution Management*, *41*(11/12), 901–916. https://doi.org/10.1108/IJRDM-01-2013-0013

Ko, E., Costello, J. P., & Taylor, C. R. (2017). What is a luxury brand? A new definition and review of the literature. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2017.08.023

Li, G., Li, G., & Kambele, Z. (2012). Luxury fashion brand consumers in China: Perceived value, fashion lifestyle, and willingness to pay. *Journal of Business Research*, *65*(10), 1516–1522. https://doi.org/10.1016/j.jbusres.2011.10.019

Liu, C., Borg, I., & Spector, P. E. (2004). Measurement equivalence of the German Job Satisfaction Survey used in a multinational organization: Implications of Schwartz's culture model. *Journal of Applied Psychology*, *89*(6), 1070.

Luxe Digital (2021). The 15 Most Popular Luxury Brands Online in 2021. Retrieved from https://luxe.digital/business/digital-luxury-ranking/most-popular-luxury-brands/.

Maziriri, E. T., & Madinga, N. W. (2018). Data to model the prognosticators of luxury consumption: A partial least squares-structural equation modelling approach (PLS-SEM). *Data in Brief*, *21*, 753–757. https://doi.org/10.1016/j.dib.2018.10.032

Mclnnis, K. (2021). The real reason Louis Vuitton and Chanel are raising their prices? Brands aren't just weathering the pandemic – luxury goods only get more desirable when they're less accessible. *South China Morning Post.* Retrieved from https://www.scmp.com/magazines/style/luxury/article/3127998/real-reason-louis-vuitton-and-chanel-are-raising-their.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111–130.

Motl, R. W., Dishman, R. K., Birnbaum, A. S., & Lytle, L. A. (2005). Longitudinal invariance of the Center for Epidemiologic Studies-Depression Scale among girls and boys in middle school. *Educational and Psychological Measurement*, *65*(1), 90–108.

Mucci, N., Giorgi, G., Roncaioli, M., Fiz Perez, J., & Arcangeli, G. (2016). The correlation between stress and economic crisis: A systematic review. *Neuropsychiatric Disease and Treatment*, 983. https://doi.org/10.2147/NDT.S98525

Mühlmann, H. (1975). *Luxus und Komfort—Wortgeschichte und Wortvergleich*.

Parguel, B., Delécolle, T., & Valette-Florence, P. (2016). How price display influences consumer luxury perceptions. *Journal of Business Research*, *69*(1), 341–348. https://doi.org/10.1016/j.jbusres.2015.08.006

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Statista Research Department. (2020). Most used tools in market research in the U.S. 2017-2018. In *Statista*. Retrieved from https://www.statista.com/statistics/917601/market-research-industry-us-most-used-tools/.

Statista Research Department. (2020). Share of English speakers by region India 2019. In *Statista*. Retrieved from https://www.statista.com/statistics/1007578/india-share-of-english-speakers-by-region/.

Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.

Stokburger-Sauer, N. E., & Teichmann, K. (2013). Is luxury just a female thing? The role of gender in luxury brand consumption. *Journal of Business Research*, *66*(7), 889–896. https://doi.org/10.1016/j.jbusres.2011.12.007

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *5*(3), 139–158.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology*, *78*(4), 557.

Vigneron, F., & Johnson, L. W. (2004). Measuring perceptions of brand luxury. *Journal of Brand Management*, *11*(6), 484–506. https://doi.org/10.1057/palgrave.bm.2540194

Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology & Marketing*, *19*(4), 357–368.

Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, *52*(1), 1–14.

Yu, S., & Hu, Y. (2020). When luxury brands meet China: The effect of localized celebrity endorsements in social media marketing. *Journal of Retailing and Consumer Services*, *54*, 102010.

Yu, S., Hudders, L., & Cauberghe, V. (2017a). Tracking the Luxury Consumer Online: An Experimental Study on the Effectiveness of Site and Search Retargeting for Luxury Brands in China and the Netherlands: An Abstract. In P. Rossi (Ed.), *Marketing at the Confluence between Entertainment and Analytics* (pp. 1541–1541). Springer International Publishing. https://doi.org/10.1007/978-3-319-47331-4_301

Yu, S., Hudders, L., & Cauberghe, V. (2017b). Targeting the luxury consumer: A vice or virtue? A cross-cultural comparison of the effectiveness of behaviorally

targeted ads. *Journal of Fashion Marketing and Management: An International Journal*, *21*(2), 187–205. https://doi.org/10.1108/JFMM-07-2016-0058

Yu, S., Hudders, L., & Cauberghe, V. (2018). Selling luxury products online: The effect of a quality label on risk perception, purchase intention and attitude toward the brand. *Journal of Electronic Commerce Research*, *19*(1), 16–35.