Effects of urban environmental attributes on graduate job preferences in Northeastern China: an application of conjoint analysis and big data methods

Jiahao Zhang, Hiroatsu Fukuda, Xindong Wei, Li Zhang* and Jinming Jiang

1 Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan
2 School of Environmental and Municipal Engineering, Jilin Jianzhu University, Changchun, People’s Republic of China
3 School of Architecture and Urban Planning, Shenzhen University, Shenzhen, People’s Republic of China
4 Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology, Qingdao, People’s Republic of China

* Author to whom any correspondence should be addressed.
E-mail: zhangli04865@gmail.com

Keywords: sustainable development, big data, job preferences, conjoint analysis

Abstract

A constant supply of novel ideas and contributions from all economic sectors is required to further the sustainable development of cities. Therefore, there is a growing need for well-educated graduates to enter metropolitan job markets. As urban environments and culture have been shown to affect a graduates’ eventual carrier choice and trajectory, governments often seek to change their local environments to attract graduates who can help efficiently allocate and utilize a city’s often-limited environmental budgets. In this study, the conjoint analysis (CA) method was employed to explore the effects of four environmental attributes (water pollution, air pollution, littering, and green area) on graduate employment preferences in northeast China. Water pollution was shown to have the greatest effect on graduate preferences (43.6%), followed by air pollution (34.1%), littering (20.7%), and green area (1.6%). According to this ranking of importance, cities could improve their environmental attributes to maximize the attraction of Northeast graduates. Moreover, this study applied the Baidu index (a big data sharing platform) to improve the attribute selection process of the CA method. The improvement reduced the cost of the CA method and enhanced its objectivity.

1. Introduction

The development of sustainable cities requires significant human resources [1]. However, it is difficult to find well-educated employees and preserve them for a long term [2]. In addition to this human capital, urban sustainability is directly affected by several other factors, particularly the local environment, economy, and society [3]. City governments can adjust these three factors to efficiently attract well-educated human resources and effectively advance sustainable development goals. Therefore, many policies have been issued by Chinese metropolitan governments that aim to attracting skilled workers.

Specifically, cities in Northeastern China have been attracting highly educated graduate workers. Northeastern China consists of three provinces (Heilongjiang, Jilin and Liaoning) with five cities in the Inner Mongolia Autonomous Region (i.e. Chifeng City Tongliao City, Xing’an League, Xilingol League and HulunBuir City). In 2011, Northeastern China accounted for 8.82%, 10.64%, and 15.11% of China’s total population, the gross domestic product and total land area respectively [4]. Due to poor economic development, there has been a greater percentage of emigration from Northeastern China than from other regions in China, with the majority of emigrants preferring to work in Chinese metropolises [5]. In addition to this increased propensity for emigration, Northeastern graduates express an increased willingness to stay and live in the cities where they began to work [6]. Combined, these traits make Northeastern graduates a stable source of long-term highly-skilled immigrants who can assist in the sustainable urban development of metropolitan areas across China [7].
In this article, the term ‘job preferences’ is used to refer to various employee tendencies regarding the process of career selection. Typically, people decide where to live and work by measuring several critical attributes and comparing them between cities. These attributes often include wealth effects [8], city size, geographical preferences of employees [9–11], social services, income, and population [12]. Previous studies have also indicated that local environmental attributes affect job preferences [13]. Indeed, many economically prominent regions have created initiatives that promote environmental sustainability and clean-up to attract well-educated talent [14]. However, while numerous studies have suggested that environmental attributes are important for attracting talent, the relative importance of different attributes has yet to be explored.

Previously, studies have sought to quantify the level or health of urban environments by combining several of the most important environmental attributes. Zenker et al. applied the conjoint analysis (CA) method to compare several German cities and analyze their attractiveness to well-educated graduates and talents [15]. Their research grouped several environmental attributes, including the number of parks, amount of pollution, and access to water, into one factor, ‘Nature and Recreation’. De Noni et al. [16] combined quantifications of a city’s green areas, air pollution, and waste management efficiency into a single parameter to assess Milan, Italy. Similarly, Merrilees et al. [17] utilized a ‘clean environment’ factor, which accounted for the cleanliness and pollution of a city’s local environment. However, not all studies clearly delineate which approach is applied to determine the most relevant environmental criteria or the rank of importance given to each attribute. This information is vital for informing future studies as well as policy-level decisions regarding the distribution of the often-limited environment-related budgets of local governments.

CA, as applied by Zenker et al. [15], first became a widely applied marketing research technique used to assess consumer multi-attribute utility functions in the 1970s [18, 19]. It has since become one of the most common approaches for exploring individual preferences over the past four decades [20, 21]. CA was effectively used to assess the non-market value of chosen attributes [20] and was initially applied to environmental studies by Beggs et al. [22]. CA has even been applied to determine the willingness of participants to pay for environmental issues [23]. Thus, due to the well-documented reliability of CA, it was chosen to examine graduate attitudes toward different urban environmental attributes in this study.

While CA can be an effective tool, studies that employ this method are generally limited regarding the number of attributes that can be considered in a single study, as a larger pool of attributes leads to an increased number of questions on the corresponding questionnaire. This can result in respondents experiencing a fatigue effect that reduces the reliability of the results [20, 24]. Moreover, research has indicated that individuals make decisions primarily by considering only a few critical attributes [25, 26]. It is possible to use the CA method to assess respondent preferences for a long list of relevant attributes; however, this requires complicated mathematical models (i.e. hybrid CA and adaptive CA) [27, 28]. Therefore, it is common to use information from preliminary experiments or consultation with experts to determine the most important parameters and reduce the number of study attributes and questions [27–29]. However, conducting exploratory experiments often costs excess time and money.

Fortunately, big data technology has been shown to accurately predict human preferences and behaviors [30, 31], creating an opportunity for a new attribute number reduction method to be explored and established in this study. The Baidu index is a big data sharing platform constructed by Baidu’s massive users’ behavior information and it represents a normalized search volume for selected keywords over a specified period [32, 33]. Since Baidu is the most commonly used search engine by Chinese people, the Baidu index has become an important source for many big data studies in China [21, 34–36]. The application of the Baidu index may help researchers to improve traditional CA methods by establishing an objective low-cost attribute number reduction process.

In this study, the Baidu index was applied to develop and optimize a new attribute reduction process for the CA method. Moreover, the significance of each selected urban environmental attribute was explored in terms of its impact on Northeastern graduate job preferences. Finally, the differences among recent graduates with different expected incomes were also examined.

The remainder of this paper is organized as follows; section 2 introduces the research method and verifies the feasibility of simplifying the conventional CA process by exploiting Baidu index. Section 3 explains the results of the CA experiments including the importance rank of the different environmental attributes regarding their influence on graduate preferences, as well as the correlation between the expected income of recent graduates and their environmental preferences. Section 4 elucidates implications of our results and compares our results with the conclusions from other similar studies. Section 5 presents the conclusions of our work and provides suggestions for future research. An outline is provided at the end of the study.

2. Methods

A face-to-face survey was designed using the Tencent questionnaire (https://wj.qq.com/). All data were
acquired between March 2019 and May 2019, and a mathematical model was built using the statistical software environment, R. An exploratory experiment examined the correlation between the Baidu index of environmental attributes and graduate preferences for the attributes.

2.1. Conjoint analysis (CA) method
CA was chosen to assess graduate job preferences in this study due to its well-documented reliability and use over time (see section 1). Relevant urban environmental attributes were chosen by the big data sharing platform of Baidu company. This method of attribute selection was chosen because it required significantly less time and money than preliminary analyses. All processes of the CA needed to be accomplished rapidly before the respondents graduated, as students usually plan their future careers and form certain psychological expectations in the months prior to graduation [37]. Furthermore, expert consultation is typically regarded as a highly subjective method that is difficult to apply uniformly to varying situations and was thus not considered in this study [38].

2.2. Sampling process
All respondents in this study were from one of five Chinese universities. At least one university was chosen from each major province in Northeastern China. According to the comprehensive strength ranking from New Oriental Education & Technology Group Inc. (Xindongfang in Chinese) [39], the largest education company in China, five universities at different levels (Jilin University (ranking 9th), Harbin Engineering University (ranking 63rd), Dalian Maritime University (ranking 112nd), Changchun University of Science and Technology (ranking 171st) and Jilin Jianzhu University (ranking 433rd)) was selected as sample sources. In 2017, 10.8% of graduates in China were from the top 50 Chinese universities [40]. Thus, we selected approximately 10.8% of respondents from Jilin University, and the remaining respondents were chosen from the four other universities. The 2017 graduate data were also used to determine the educational level of respondents, with 85% graduating with a bachelor’s degree, 13.5% with a master’s degree (13.5%) and 1.5% with a doctoral degree.

We planned to recruit 500 respondents for the preliminary experiment and 1600 respondents for the main experiment. Through cooperation with the staff of the five universities, we randomly selected respondents in the graduating class according to the proportions above and distributed questionnaires to their mobile phones or computers. Finally, 483 valid questionnaires were collected for the preliminary experiment, and 1589 valid questionnaires were collected for the CA experiment.

2.3. Experimental process design
As shown in figure 1, the urban environmental attributes and levels required to be considered for this experiment were identified in Step 1. A preliminary experiment was utilized to select these attributes. In the preliminary experiment, we selected the most relevant attributes using the traditional questionnaire method alongside the new Baidu index selection method. The results of the two selection methods were compared to verify the reliability of the new method. In Step 2, the main-experiment questionnaire was designed to examine the characteristics and socioeconomic information of respondents and their preferences for the combinations of different urban environmental attributes and levels. Table 1 outlines the respondent’s reported demographics. We used the orthogonal processing function of SPSS software to design these attributes and levels into nine combinations. To examine graduate preferences, these were portrayed as nine virtual cities that the respondents scored in the CA portion of the questionnaire (see appendix 1). Then, in Step 3, the questionnaire data were analyzed based on the model established using R version 3.4.3 [41], and the part-worth utility (PWU) and relative importance results were obtained. Finally, in Step 4, we evaluated these results to identify overall trends in graduate preferences for urban environmental attributes and job preferences of graduates with different expected incomes.

2.4. Preliminary experiment
A preliminary experiment was conducted to rank the urban environmental attributes based on their importance using two methods. The first method was based on the Baidu index. The Baidu index...
Table 1. Four urban environmental attributes included in the conventional CA task. (P is the proportion of annual air pollution days and A is park green area per capita.).

| Attributes   | Attribute definition                                                                 | Attribute levels      |
|--------------|--------------------------------------------------------------------------------------|-----------------------|
| Air pollution| In this study, the attribute of air pollution is quantified as the frequency of occurrence of fine particulate matter (PM 2.5) pollution in a year. | $P < 5\%$            |
|              |                                                                                      | $5\% < P < 20\%$     |
|              |                                                                                      | $20\% < P < 35\%$   |
| Water pollution| The attribute of water pollution is defined as the contamination of water sources in city. | No                    |
| Littering    | The attribute of littering refers to the existence of garbage in areas such as citizens’ living quarters, workplaces, and major transportation routes. | No                    |
| Green area   | Green area refers to the size of the vegetation coverage area in the main living area of the citizens. In this study, this attribute is quantified by the city’s park green area per capita. | $5 \text{ m}^2 < A$   |
|              |                                                                                      | $10 \text{ m}^2 < A < 20 \text{ m}^2$   |
|              |                                                                                      | $20 \text{ m}^2 < A$   |

Figure 2. Monthly Baidu index of various metropolis environmental attributes in 2018.

Figure 3. Comparison between Baidu index and pre-test score of environmental attributes.

refers to the total number of times a particular word was searched on the Baidu website within a period specified by the researchers. On the Baidu website, researchers can select a specific word and determine the Baidu index of other words that are most closely related to this word. Through this function, we obtained the Baidu index for all attributes related to the urban environment in 2018. There were seven urban environmental attributes identified with an average of 10 or more searches per day: air pollution, water pollution, green area, littering, noise pollution, soil pollution, and light pollution. Figure 2 presents the monthly Baidu index of these attributes for the entire period of 2018. A higher Baidu index indicates more searches. Objects with more searches have attracted more interest from people [42, 43], meaning that they are perceived as more important than other objects with fewer searches.

To examine the similarity between the urban environmental attribute ranks obtained with the Baidu index and those obtained with the traditional questionnaire method, we used web-based questionnaires to explore graduate preferences for the seven attributes mentioned above. They were asked to score them from 7 to 1 (7-point Likert scale), where higher scores represented greater importance. Graduate preferences regarding urban environmental attributes were then determined using the total scores from 483 respondents. This was compared with the results from the Baidu index analysis (figure 3).

A similar trend was found between the Baidu index and pre-test score. SPSS software was used to examine the Spearman correlation coefficient, which was 0.964 ($p < 0.001, N = 7$), suggesting that the Baidu index was highly correlated with the pre-test scores. Thus, to avoid the fatigue effect and focus only on the most important attributes, we selected the four environmental attributes with the highest Baidu index for the CA: air quality, water pollution, littering, and green area.

There exist several CA methods including conventional CA, hybrid CA, adaptive CA and hierarchical CA [44]. To achieve high performance in data analysis with less than five attributes, the conventional CA method was chosen for the following experiment [20].

2.5. Conjoint analysis (CA) process

The survey was conducted in 10–15 min of class time, as assisted by university staff. The first part of the questionnaire focused on respondent characteristics
lists the demographic characteristics (i.e., gender, age, educational background, income, residential area).

The second part was the CA, where attribute levels were set according to the standards established by the Chinese government. In China, urban air quality is usually assessed using the annual percentage of days where air pollution is higher than a set threshold. Green area was assessed as the amount of park green area per capita. The levels of these two environmental attributes were set using data from the four most developed Chinese cities (Beijing, Shanghai, Guangzhou, and Shenzhen) [45]. The levels of the other two attributes (water pollution and littering) directly refer to the pollution degree classification of the Chinese urban environment [46, 47]. Before answering the questionnaire, respondents were confirmed to be able to understand and distinguish each study attribute and level.

Orthogonal processing was used to simplify the levels of each attribute into nine different combinations. Table 1 lists the selected urban environmental attributes along with their levels. Bigsby and Ozanne [48] found that visual stimuli can help respondents understand the different attribute and level combinations; thus, respondents were presented with pictures to better understand different environmental attributes.

Additionally, participants were asked to respond to a traditional CA questionnaire. This model was first described by Louviere and Woodworth [49]. Shocker and Srinivasan further suggested that the attributes of CA method should be made operational [42]. Consequently, the respondents were required to indicate if they had previous experience with poor levels of any of the four study attributes. Then, the staff guided the respondents to recall their experience, personally involving them in the simulated situation. Then, respondents were required to rate the nine simulated pictures of the different attribute and level combinations. The rating scale ranged from 0 to 10 where 0, suggested that working in that city was unacceptable, whereas a ten indicated that it was extremely desirable to work there.

2.6. Mathematical model and data processing

The findings of the CA section of the survey were analyzed for all samples and 11 different social-demographic and personal variables were additionally accounted for. Graduate preferences were calculated from the results of the CA section using a multinomial logit function. The function determined the importance of the respective attribute on the graduates’ decision making relative to the other attributes, as well as the level of importance of each level of the attributes (PWU). The derived graduate preference model was thus expressed using:

\[ U = \beta_0 + \sum_{k=1}^{n} \beta_k X_k \]  

where \( \beta_0 \) denotes a constant coefficient for each alternative, and \( \beta_1, \beta_2, \beta_3, \ldots, \beta_n \) represent the coefficients obtained by the logit model and indicate the relative importance of the attributes in each alternative. The relative importance of attributes suggests their importance for decision making and the preferences for all levels in the attribute. The relative importance values of each attribute were calculated by determining the PWUs for each level of the attribute. PWU values show how much a specific level of an attribute is desired or unwanted. For example, the ‘water pollution’ attribute has three levels (no pollution, slight pollution, and serious pollution). If no pollution, slight pollution, and serious pollution had PWU values of 7, 0, and −4, respectively, then the ‘unpolluted’ level is the most desired level, with a positive PWU. ‘Slight pollution’ is neither desired nor unwanted, and ‘serious pollution’ is unwanted. The relative importance of an attribute is the ratio of the difference between its highest and the lowest PWU to the sum of the differences between the highest and lowest PWU of all attributes.

Moreover, we adopted the demographics hierarchical Bayes model to analyze the 1589 samples of the main study. The model was built using R's Bayesm package [50] for the hierarchical Bayes random-effects model, which is capable of estimating general and individual parameters simultaneously.

3. Results

Table 2 lists the demographic characteristics of the respondents.

3.1. Part-worth utilities (PWUs)

The PWUs of all attribute levels increased with decreasing pollution, as expected. However, ‘green area’ did not increase as significantly as the other attributes (figure 4). The PWUs of this attribute's levels were extremely close, indicating that all levels were relatively desirable for most respondents. For air pollution, an attribute level of less than 5% was significantly more desirable (PWU = 2.672) than the 5%–20% level (PWU = 1.346) or the 20%–35% level (PWU = −0.2731). ‘Water Pollution’ exhibited the most significant PWU difference of all tested attributes. The PWUs of the three levels (None, Slight, Serious) were 3.127, 1.682 and −0.637, respectively, suggesting that respondents prefer no water pollution almost twice as much as slight water pollution. Moreover, the PWU values indicated that ‘no littering’ had similar desirability to ‘slight littering,’ and both of these less-severe levels were significantly more preferred than ‘serious littering.’
Table 2. Characteristics of participants (n = 1589).

| Items                        | Overall |
|------------------------------|---------|
| Age                          |         |
| <20                          | 20      |
| 20 ～ 25                     | 70.3    |
| >25                          | 9.7     |
| Sex                          |         |
| Female                       | 44.542  |
| Male                         | 55.173  |
| Marriage status              |         |
| Unmarried                    | 96.9    |
| Married                      | 2.8     |
| Education                    |         |
| Below bachelor’s degree      | 4.4     |
| Bachelor’s degree            | 79.8    |
| Above bachelor’s degree      | 15.8    |
| Monthly salary (CNY)         |         |
| <1000                        | 15.8    |
| 1000 ～ 2000                 | 55.9    |
| 2000 ～ 5000                 | 10.8    |
| >5000                        | 4       |
| Inconvenient                 | 13.2    |
| Hometown                     |         |
| Urban                        | 54.8    |
| Rural                        | 45.2    |
| Parents’ education           |         |
| Below bachelor’s degree      | 71      |
| Bachelor’s degree            | 26.6    |
| Above bachelor’s degree      | 2.4     |
| Annual household income (CNY)|         |
| <30 000                      | 18.6    |
| 30 000 ～ 50 000              | 15.8    |
| 50 000 ～ 100 000             | 18      |
| 100 000 ～ 200 000            | 15.2    |
| >200 000                     | 5.9     |
| Inconvenient                 | 26.2    |
| Willingness to work in       |         |
| metropolitan areas           |         |
| Yes                          | 65.1    |
| No                           | 34.9    |
| Air pollution experience     |         |
| Yes                          | 88.5    |
| No                           | 11.6    |
| Water pollution experience   |         |
| Yes                          | 74.4    |
| No                           | 25.6    |
| Expected future monthly      |         |
| income (CNY)                 |         |
| <3000                        | 2.1     |
| 3000 ～ 5000                 | 8.3     |
| 5000 ～ 8000                 | 35.2    |
| 8000 ～ 15 000               | 42.2    |
| >15 000                      | 11.9    |

3.2. Relative importance

Figure 5 presents the relative importance of the four selected attributes. The results revealed that water pollution (43.6%) was the most important attribute, followed by air pollution (34.1%), littering (20.7%), and green area (1.6%). The relative importance of Per capita green area was significantly lower than the other attributes. The rank of their relative importance is completely identical to the rank of their Baidu indexes (figure 2).

Figure 6 shows that the relative importance of urban environmental attributes changed with an increase in expected future income. To avoid the effect of education level, only students graduating with a bachelor’s degree were selected for this analysis. The
results showed that the importance of littering and air pollution increased with expected future income, the importance of water pollution decreased, and the importance of green areas was the greatest for graduates with the lowest expected future income.

4. Discussion

For the four urban environmental attributes examined in this study (water pollution, air pollution, littering, and green area), our results suggest that similar trends existed for the Baidu index and pre-test score results. Combined with a high Spearman correlation (0.964), our validation indicates that environmental attributes for the CA method can be accurately selected using the Baidu index, as both preliminary analyses created similar rankings. This result means that a Baidu index-based attribute selecting method can replace the traditional questionnaire method, solving the problem of long experimental time and high expenses while still retaining its utility. In the study by Arning et al that investigated the public acceptance of sustainable CO$_2$-derived building materials, the preliminary experiment used five participants (out of 145 total participants) to determine the attributes for the main experiment [51]. In a study by Chelsea et al, ten patients (out of 200 total participants) were selected for one-to-one 20 min interviews that investigated the most important attributes of arthrodesis and arthroplasty in their cognition [52]. For studies with too few pre-experimental participants, use of the Baidu index method to filter attributes can help to avoid subjective results due to a shortage of interviewees. Additionally, CA of environmental cases often requires a large number of samples, so pre-experiments also require a large number of respondents to make them effective. Gil-Hwan Lim adopted a joint analysis method to investigate South Korean residents’ attitudes towards nuclear power plants [53]. Researchers used the participants in the preliminary experiment used their questionnaire responses of 81 participants (out of a total of 1417 participants) in the preliminary experiment to identify the three most important attributes for all respondents. For CA experiments in the environmental field, the Baidu index method can be used to shorten the experimental period by simplifying the pre-experiment and greatly reducing experiment costs. The Baidu index is just a basic application of big data technology on the Chinese search engine platform. There are still many limitations to its application, particularly the lack of users in English-speaking regions of the world. Compared with Baidu index, Google Trends is a search behavior analysis platform with similar functions for English users. It is believed that Google Trends can replace the role of Baidu index in the attribute selection process of CA in order to apply this method on a more global scale.

Our results indicate that the most important environmental attribute influencing graduate job preferences is water pollution, followed by air quality, littering, and green area. These results indicate that improvements to water pollution issues will be more attractive for Northeast graduates than improvements to other environmental attributes. For graduates from other regions, the results of this study may not be applicable. Previous studies have suggested that green space and cleanliness are the most important environmental attributes for attracting workers or students [54, 55]. In a study evaluating the preferences of residents in Porto, Portugal, the most important environmental attribute was green space, which was much more important than urban cleanliness or pollution (air, water, and noise) [56]. Conversely, environmental quality (green areas, urban waste management, and noise/air pollution prevention) was shown to have a more negative impact on the attractiveness of Milan, Italy to its residents [16]. It is likely that regional differences have caused the disparities between the results of our study and those of other cities and regions. Graduates from Northeastern China have different preferences and experiences than respondents from the Americas or Europe.

For developing Chinese cities, Kumar et al indicates that the quality of water resource plays a critical role in their development, and preventing water pollution is the most important part of urban environmental pollution control [57]. These results show similarities to the environmental preferences of Northeast graduates in our study.

Besides Beijing, Shanghai, Guangzhou and Shenzhen, many fast-growing Chinese cities also have a huge demand for talents. Due to the significant differences in scale between these metropolises and other Chinese cities, it is necessary to minimize the gap between non-environment factors to better reflect the differences in the attractiveness of different environmental factors. Based on the agglomeration of commercial resources, urban hubs, urban resident activity, lifestyle diversity, and future plasticity levels, Chinese cities are divided into six levels. The number of cities at different levels is as follows: 4 first-tier cities, 15 new first-tier cities, 30 second-tier cities, 70 third-tier cities, 90 fourth-tier cities, 128 fifth-tier cities [58]. It is feasible to select cities of the same level and adopt our method to evaluate the environmental attractiveness to graduates. However, our research still has other limitations. Although our results point out that improving water pollution is the most important environmental measure, the graduates’ detailed preferences of graduates on ways to improve water quality and their willingness to pay still need to be examined.

Expected future income refers to income estimated based on the past experiences of an individuals. Graduates with different expected future incomes varied in their indicated varying preferences for the
environmental attribute levels. The graduates expecting higher incomes tended to select cities with higher levels of positive environmental attributes (e.g., less air pollution and littering). However, graduates expecting lower income were more concerned about local urban water pollution. Green areas did not exhibit any correlation with the expected future income of graduates. Since graduates usually have no real income before they graduate, we classified respondents based on their expected future income and examined the subsequent differences in their urban environmental preferences. The reason for the difference in preferences of students with different expected incomes might be that graduates with higher expected incomes have enough budgeted to select a living place which is surrounded by a clean environment. Air quality is not particularly different for various regions in the same city, thus ‘air pollution’ was the urban environmental attribute they considered firstly. Jacobsen, Lundhede and Thorsen indicated that the expected future is critical to environmental attitudes [59]. A study of 39 cities in the Czech Republic documented that inhabitants with unfavorable socioeconomic status mainly reside in smaller cities with higher concentration levels of combustion-related air pollutants [60]. A study by Liu et al. revealed that the increase in air pollution concentration has short-term positive production effects but long-term negative impacts on individual income and will exacerbating the income inequality across socioeconomic statuses [61]. We speculate that graduates with higher expected incomes may have relatively strong confidence in their careers and are unwilling to trade long-term negative effects on their careers for short-term positive production effects, so they choose living place with clean environment. Li et al. also indicated that regional environmental pollution will widen the income gap of local workers [62]. Regional income inequality is likely to affect the tendency of graduates to choose employment in this place. Future research on this subject could apply expected income as a special social background factor in urban environmental research questionnaires to classify unpaid people and obtain their different preferences. A study by Wang et al. showed that absolute income has a positive correlations with environmental concerns [63]. Li and Chen argued in their study that relative income is more likely to exert a significant effect on environmental concerns than that of absolute income [64]. However, most of graduates do not yet have an income. Instead, their welfare is still dependent on their expected future income [65].

While graduates did indicate that it was undesirable to live in a location with a high level of littering, they overwhelmingly suggested that air and water pollution were the most critical environmental attributes that were used to help determine the location of their first carrier job. Thus, if policymakers focus only on improving these two attributes, it should be effective in attracting graduates to urban centers. However, the attribute of ‘green area’ should not be considered a necessary attribute to promote, as our results showed no significant differences in preferences for the different levels. Therefore, campaigns to increase green areas will not make any significant impact on skilled-worker immigration, meaning urban governments should focus on other environmental problems instead of increasing their green areas. Additionally, the close PWU values found for the low and moderate levels of ‘littering’ indicate that graduate preferences will not significantly change if a city with ‘slight littering’ enacts campaigns to reach ‘no littering’. Thus, governments should not necessarily strive to improve littering from ‘slight’ to ‘none’, as they will receive no added benefit in skilled labor influx.

5. Conclusions and suggestions

In this study, we utilized the CA method to examine the importance of various urban environmental attributes for attracting graduates to work in metropolises. This paper employed the Baidu index in a novel manner to improve the attribute selection process of the CA method, reducing its cost and enhancing its objectivity.

Data on graduate job preferences were collected via questionnaires. Then, a mathematical model was created within R’s statistical environment to analyze the obtained data. The primary conclusions of this study are as follows:

(a) The Baidu index can be feasibly and efficiently employed in the attribute selection process of the CA method.
(b) Water pollution was the most critical urban environmental attribute with the highest importance (43.6%), followed by air pollution (34.1%), littering (20.7%), and the green area (1.6%).
(c) Local governments should focus on improving local water and air quality to attract highly-educated graduate workers; littering should simply be maintained at or below a moderate level.
(d) Graduates with higher expected future income prefer to address the issues of air pollution and littering over water pollution. Conversely, graduates with lower expected future incomes prefer addressing water pollution over any other environmental attributes of metropolises.

In this study, we discussed which environmental attribute levels are preferred by graduates. This information can be used to help inform the spending decisions of regional or city governments to optimize the most crucial attributes. It should be noted,
however, that the eventual profit of attracting graduates optimizing the local environment to their preferences depends greatly on the methods of environmental improvement and how the local budget is divided.

In the future, several problems still require further examination, primarily including the following:

(a) Graduates with different socioeconomic information will have different preferences for various combinations of the urban attributes. Researchers should examine how different socioeconomic backgrounds could influence the graduate preferences in selecting the city of their first career job.

(b) Research should seek to broaden the applications of the Baidu index in other fields related to the construct of cognitive.

Data availability statement
The data that support the findings of this study are available upon reasonable request from the authors.

Funding
This research is sponsored by National Natural Science Foundation of China (Grant No. 51478268) and Chinese Postdoctoral Science Foundation (2020M682886).

Acknowledgments
I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to Professor Hiroatsu Fukuda, my supervisor, for his constant support and guidance. I also owe a special debt of gratitude to the staffs in Luoyang tourist attractions. They have provided the necessary support for my work.

Conflict of interest
The authors declare no conflict of interest.

Appendix 1

Questionnaire for main experiment
We are conducting a survey which examines graduates’ preferences on the metropolises with different environmental attributes’ levels. We sincerely hope that you can spend a few minutes on this survey, and we would definitely value your opinions! Your identity and answers are absolutely confidential!

(a) Age
   <20
   20–25
   >25
(b) Gender
   Female
   Male
(c) Marital status
   Unmarried
   Married
(d) Highest academic qualification
   Below bachelor’s degree
   Bachelor’s degree
   Above bachelor’s degree
(e) Monthly income or living funds (CNY)
   <1000
   1000–2000
   2000–5000
   >5000
   Inconvenient
(f) Residence
   Urban
   Rural
(g) Highest academic qualification of your parents
   Below bachelor’s degree
   Bachelor’s degree
   Above bachelor’s degree
(h) Family income per year (CNY)
   <30 000
   30 000–50 000
   50 000–100 000
   100 000–200 000
   >200 000
   Inconvenient
(i) Have you considered working in metropolises in the past?
   Yes
   No
(j) Which of the following environmental pollution situations have you experienced? ()
   Water pollution
   Air pollution
(k) If you are going to work in the metropolis, what is your expected monthly income? (CNY)
   <3000
   3000–5000
   5000–8000
   8000–15 000
   >15 000

In order to helping you understand the different levels of environment attributes in our questionnaire clearly, we will show you some related pictures (see figure A1–A9).

In this survey, we set three levels for four environmental attributes (air pollution, water pollution, littering, green area) respectively, and nine virtual metropolises were combined with different levels of these attributes (detailed levels refer to table A1).

You need to evaluate your willingness to work in these nine virtual metropolises and score on a scale 1–10 (1 represents absolutely rejection, 10 is
Table A1. Four urban environmental attributes included in the conventional CA task. (P is the proportion of annual air pollution days and A is park green area per capita.)

| Attributes      | Attribute levels                        |
|-----------------|-----------------------------------------|
| Air pollution   | P < 5%                                   |
|                 | 5% < P < 20%                             |
|                 | 20% < P < 35%                            |
| Water pollution | No                                      |
|                 | Slight                                   |
|                 | Serious                                  |
| Littering       | No                                      |
|                 | Slight                                   |
|                 | Serious                                  |
| Green area      | 5 m² < A                                 |
|                 | 10 m² < A < 20 m²                         |
|                 | 20 m² < A                                |

extremely eager). Please carefully weigh these nine virtual metropolises and score them according to your will.

(a) Annual air pollution days proportion < 5%, slight level of water pollution, serious level of littering, per capita park green area > 20 m²;
(b) Annual air pollution days proportion < 5%, serious level of water pollution, no littering, 10 m² < per capita park green area < 20 m²;
(c) 5% < Annual air pollution days proportion < 20%, no water pollution, slight level of littering, 10 m² < per capita park green area < 20 m²;
(d) 5% < Annual air pollution days proportion < 20%, serious level of water pollution, slight level of littering, per capita park green area > 20 m²;
(e) 5% < Annual air pollution days proportion < 20%, slight level of water pollution, no littering, 5 m² < per capita park green area < 10 m²;
(f) 20% < Annual air pollution days proportion < 35%, serious level of water pollution, serious level of littering, 5 m² < per capita park green area < 10 m²;
Thank you for taking the time to complete this questionnaire for us!

ORCID ID

Jiahao Zhang https://orcid.org/0000-0002-3331-3092

References

[1] Suganthi L 2018 Multi expert and multi criteria evaluation of sectoral investments for sustainable development: an integrated fuzzy AHP, VIKOR/DEA methodology Sustain. Cities Soc. 43 141–56
[2] Vauterin J J, Michelsen K E and Linnanen L 2013 Attracting and retaining student talent from around the world: the lived experience in university–industry collaboration Ind. Higher Educ. 27 163–78
[3] Li W and Yi P 2020 Assessment of city sustainability—coupling coordinated development among economy, society and environment J. Clean. Prod. 256 120453
[4] National Bureau of Statistics of the People’s Republic of China 2013 China Economic Census Yearbook 2013 (Available at: www.stats.gov.cn/tjsj/ycsj/jyjc/3jpy/indexhx.html) (Accessed 19 June 2019)
[5] Zhao F and Liu Y 2018 A study on the causes of population loss in northeast china—based on time-varying coefficient factor model Popul. J. 40 81–90
[6] Xu L 2021 An analysis of the factors affecting the employment of college graduates in the province and countrimeasures: taking the school of humanities of northeast normal university as an example China Univ. Stud. Career Guide 4 52–55
[7] Jianyu Z and Yu G 2018 The loss of highly educated population in Northeast China and its reasons—based on the analysis of graduate employment data of Jilin University from 2013 to 2017 Popul. J. 40 55–65
[8] Haywood L 2016 Wealth effects on job preferences Labour Econ. 38 1–11
[9] Chowdhary M et al 2019 The impact of graduates’ job preferences on the current radiation oncology job market Int. J. Radiat. Oncol. Biol. Phys. 104 27–32
[10] Gilman M, Pit V, Batyueva M and Sumik E 2020 Which cities do we like to live in? Empirical analysis of employees’ attitude to cities J. New Econ. Assoc. 45 111–30
[11] Schade M, Pieler R and Burmann C 2019 How cities can attract highly skilled workers as residents: the impact of city brand benefits on highly skilled, potential residents’ city brand attitudes: an abstract 2018 Academy of Marketing Science (AMS) World Marketing Congress (WMC) (London: Springer) p 399
[12] Glavac S M, Vias A C and Mulligan G F 1998 Population and employment interactions in the growth of United States micropolitan centers Urban Geogr. 19 632–56
[13] Plane D A, Henrie C J and Perry M J 2005 Migration up and down the urban hierarchy and across the life course Proc. Natl. Acad. Sci. 102 15313–8
[14] Florida R 2000 Competing in the age of talent: environment, amenities, and the new economy Report prepared for the RK Mellon Foundation, Heinz Endowments, and Sustainable Pittsburgh vol 60
[15] Zenker S, Eggers F and Forsky M 2013 Putting a price tag on cities: insights into the competitive environment of places Cities 30 133–9
[16] de Noni L, Orsi L and Zanderighi L 2014 Attributes of Milan influencing city brand attractiveness J. Dest. Mark. Manage. 3 218–26
[17] Merrilees B, Miller D and Herington C 2013 City branding: a facilitating framework for stressed satellite cities J. Bus. Res. 66 37–44
[18] Green P E and Rao V R 1971 Conjoint measurement-for quantifying judgmental data J. Mark. Res. 8 355–63
[19] Johnson R M 1974 Trade-off analysis of consumer values J. Mark. Res. 11 121–7
[20] Alriksson S and Öberg T 2008 Conjoint analysis for environmental evaluation Environ. Sci. Pollut. Res. 15 244–57
[21] Fang J, Gogozor G, Lau C K M and Lu Z 2020 The impact of Baidu index sentiment on the volatility of China’s stock markets Fin. Res. Lett. 32 101099
[22] Beggs S, Cardell S and Hausman J 1981 Assessing the potential demand for electric cars J. Econom. 17 1–9
[23] Shen Z, Wakita K, Oishi T, Yagi N, Kurokura H, Blasiak R and Furuya K 2015 Willingness to pay for ecosystem services of open oceans by choice-based conjoint analysis: a case study of Japanese residents Ocean Coast Manage. 103 1–8
[24] Gao S, Smits M and Wang C 2016 A conjoint analysis of corporate preferences for the sectoral crediting mechanism: a case study of Shanxi Province in China J. Clean. Prod. 131 259–69
[25] Klahr D 1970 A study of consumers’ cognitive structure for cigarette brands J. Bus. 43 190–204
[26] Raz C, Piper D, Haller R, Nicod H, Dusart N and Giboreau A 2008 From sensory marketing to sensory design: how to drive formulation using consumers’ input Food Qual. Prefer. 19 719–26
[27] Goldberg S M, Green P E and Wind Y 1984 Conjoint analysis of price premiums for hotel amenities J. Bus. 37 5111–32
[28] Johnson F R and Desvousges W H 1997 Estimating stated employment interactions in the growth of United States Econ. Educ. 163–78
[29] Alex. Eng. J. 45 163–78
[30] Veitch J, Salmon J, Deforche B, Ghekiere A, van Cauwenbergh I, Bangay S and Timperio A 2017 Park attributes that encourage park visitation among adolescents: a conjoint analysis Landscape Urban Plann. 161 52–58
[31] Viloria A, Gonzalez I R P and Lezama O P B 2019 Learning style preferences of college students using big data Procedia Comput. Sci. 160 461–6
[32] Weaver A 2021 Tourism, big data, and a crisis of analysis Ann. Tour. Res. 88 103138
[33] Li W, Yang G and Li X 2021 Correlation between PM2.5 pollution and its public concern in China: evidence from Baidu index J. Clean. Prod. 293 126091
[34] Wang T, Xia Q, Chen X and Jin X 2020 Use of Baidu index to track Chinese online behavior and interest in kidney stones Risk Manage. Healthcare Policy 13 705
[35] Yao L, Ma R and Wang H 2021 Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving average Alex. Eng. J. 60 365–72
[36] Huang X, Zhang L and Ding Y 2017 The Baidu index: uses in predicting tourism flows—a case study of the Forbidden City Tour. Manage. 58 301–6
[37] Li S, Chen T, Wang L and Ming C 2018 Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index Tour. Manage. 68 116–26
[37] Shocker A D and Srinivasan V 1974 A consumer-based methodological approach to the identification of new product ideas Manage. Sci. 20 921–37
[38] Saks A M and Ashforth B E 2000 Change in job search behaviors and employment outcomes J. Vocat. Behav. 56 277–87
[39] China University Rankings 2017 China university evaluation research report 2017 (Available at: http://gaokao.csdn.cn/201712/10746429.html) (Accessed 19 August 2021)
[40] Yearbook, C.S.I.C.S.Y. 2017 Number of Students of Formal Education by Type and Level National Bureau of Statistics of China 21–2
[41] R Core Team 2020 R: a language and environment for statistical computing (Available at: www.r-project.org/ (accessed 14 June 2020))
[42] Da Z, Engelberg J and Gao P 2011 In search of attention J. Finance 66 1461–99
[43] Preis T, Moat H S and Stanley H E 2013 Quantifying trading behavior in financial markets using Google Trends Sci. Rep. 3 1–6
[44] Poortinga W, Steg L, Vlek C and Wiersma G 2003 Household preferences for energy-saving measures: a conjoint analysis J. Econ. Psychol. 24 49–64
[45] GaWC 2018 The world according to GaWC 2012
[46] State Environmental Protection Administration 2002 [N] GB 3838–2002, Environmental Quality Standards for Surface Water[S] (Beijing: China Environmental Press)
[47] Ministry of Environmental Protection 2012 [N] GB 3095–2012, Ambient Air Quality Standards[S] (Beijing: China Environmental Press)
[48] Bigsby H and Ozanne L K 2002 The purchase decision: consumers and environmentally certified wood products For. Prod. J. 52 100–5
[49] Louviere J J and Woodworth G 1983 Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data J. Mark. Res. 20 350–67
[50] Rossi P E, Allenby G M and McCulloch R 2012 Bayesian Statistics and Marketing (New York: Wiley)
[51] Arning K, Offermann-van Heek J and Ziefle M 2021 What drives public acceptance of sustainable CO2-derived building materials? A conjoint-analysis of eco-benefits vs. health concerns Renew. Sustain. Energy Rev. 144 110873
[52] Harris C A, Shauver M J, Yuan F, Nasser J and Chung K C 2018 Understanding patient preferences in proximal interphalangeal joint surgery for osteoarthritis: a conjoint analysis J. Hand Surg. 43 615–24
[53] Lim G-H, Jung W-J, Kim T-H and Lee S-Y 2017 The cognitive and economic value of a nuclear power plant in Korea Nucl. Eng. Technol. 49 609–20
[54] Darchen S and Tremblay D G 2010 What attracts and retains knowledge workers/students: the quality of place or career opportunities? The cases of Montreal and Ottawa Cities 27 223–33
[55] Pilati T and Tremblay D G 2007 Montreal’s socio-economic development: The creative city and the artistic career as pull factors? Can. J. Reg. Sci. 30 475–95 (https://ids.ca/images/rcst/archives/V30N3-PILATI-TREMBLAY.pdf)
[56] Santos L D, Martins I and Brito P 2007 Measuring subjective health concerns building materials? A conjoint-analysis of eco-benefits vs. driving public acceptance of sustainable CO2-derived building materials? A conjoint-analysis of eco-benefits vs. health concerns
[57] Kumar P and Saroj D P 2014 Water–energy–pollution nexus China Econ. Rev.
[58] Hirschman A O and Rothschild M 1973 The changing development: with a mathematical appendix Q. J. Econ. 87 544–66