ISSN 1648-4142 / eISSN 1648-3480

VILNIUS TECH Vilnius Gediminas Technical University

TRANSPORT

2024
Volume 39
Issue 2
Pages 197–208
https://doi.org/10.3846/transport.2024.20518

Original Article

EXPLORING POTENTIAL CAR TRIPS FOR LONG-DISTANCE SCHOOL ESCORTING USING SMART CARD DATA AND A HOUSEHOLD TRAVEL SURVEY

Yang LIU[1], Yanjie JI[2✉], Xinwei MA[3], Qiyang LIU[4]

[1] *Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, China*
[2] *School of Transportation, Southeast University, Nanjing, China*
[3] *School of Civil and Transportation Engineering, Hebei University of Technology, Tianjin, China*
[4] *Peking University Shenzhen Graduate School, Shenzhen, China*

Highlights:
- propose a method for identifying school metro commuters that considers the behaviour of students who are escorted;
- use smart card data and household travel surveys to analyse potential car trips for long-distance school escorting;
- identify three usage patterns among metro-commuting students: one-way to school, one-way from school, and round-trip;
- metro usage patterns, entry times, travel durations, and the school–housing relationship significantly affect the frequency of student metro use;
- students who use the metro for round-trip commuting are often frequent riders, whereas those who use the metro for one-way trips tend to ride less frequently.

Article History:
- submitted 24 October 2017;
- resubmitted 23 February 2018, 14 December 2018;
- accepted 1 March 2019.

Abstract. Encouraging students to commute by the metro can effectively reduce household car use caused by long-distance commuting to school. This article focuses on the frequency of metro use by groups of students commuting to school based on the assumption that students who use the metro may occasionally be driven to school by their parents. For the 1st time, we propose a school metro commuter identification process that considers the potential behaviour of escorted students, and we study the potential car trips for long-distance school escorting in Nanjing (China) using Smart Card Data (SCD) and a household travel survey from Nanjing. 3 clusters of students who use the metro for commutes to school are identified by frequency of use for possible escorting behaviour based on the commuting day. As possible factors influencing the 3 frequency groups, usage pattern of the metro, entry time, travel duration and the school–housing relationship are extracted from SCD. Furthermore, a multinomial logistic regression model is used to examine the significant factors that influence the grouping of students. The results show that students who use the metro occasionally for a long commuting distance to school are more likely to be escorted to and from school by their parents, especially to school. The later the entry time is to the metro, the more likely that students are to be escorted to school. Additionally, a long school–housing travel duration/distance significantly contributes to parents' car trips for commuting. The results of this article are valuable for transport policy to reduce car use for long-distance school trips.

Keywords: transport management, school commuting, escort, smart card data, discrete choice models.

✉Corresponding author. E-mail: *jiyanjie@seu.edu.cn*

Notations

CNY – Chinese yuan;
NOCD – number of commuting days;
NOUD – number of using metro days;
SC – smart card;
SCD – SC data.

1. Introduction

Due to the increasing number of car trips by parents to escort their children to school, school travel has aroused scholars' attention in recent years (McDonald 2008a; Wilson *et al.* 2010; Li, Zhao 2015; Zhang *et al.* 2017; Mitra *et al.* 2010; Lingaitis *et al.* 2003). The increasing rates of driving students to/from school have led to not only a notable increase in traffic congestion at peak times and schoolchildren' s risk of traffic accidents but also an increase in rates of childhood obesity (Zhang *et al.* 2017; Frank, Engelke 2007; Lu *et al.* 2017). Thus, most scholars in this field have focused on ways to encourage students' independent mobility through walking or cycling (Mandic *et al.* 2017; Wen *et al.* 2008; McDonald 2008a, 2008b). The available evidence indicates that the use of cars for school commuting is closely related to the distance between the home and the school (Li, Zhao 2015; He, Giuliano 2018; Liu *et al.* 2017). A long commuting distance to school, which is usually caused by school choice in pursuit of quality educational resources, is a key factor in the increase in the use of cars. However, it is inevitable that some students will have to travel a long-distance to school.

Many developed countries have adopted the public policy of allowing freedom of school choice (Wilson *et al.* 2010). For example, such a policy resulted in only 51.3% of students enrolling in the closest school to their home in Dunedin (New Zealand) due to the demand for good educational resources (Mandic *et al.* 2017). Similar results can also be found in other countries. In many developing countries, including China, the government requires students in primary school and junior high school to follow a school enrolment policy (similar to school zoning policies in some developed countries). Nevertheless, an increasing number of parents still wish to obtain admission to high-quality schools for their children, even if it necessitates paying additional entrance fees (Zhang *et al.* 2017). These schools, however, are sometimes far from their residences. As a result, students, especially junior high school and high school students, often have to travel long-distances to and from school, and rates of active transport to school are reduced with increases in school commuting distance. Therefore, to reduce the number of car trips to school and promote behavioural changes, those who have a long school commuting distance should be a particular focus of researchers (Easton, Ferrari 2015; Li, Zhao 2015). This research will have significant implications for understanding the determinants of such travel behaviour and allow an adaptive transport policy to be proposed to encourage the use of public transit.

For medium and long-distance school commutes, urban metro transit is a good choice for students, especially in cities that lack school buses. Additionally, walking or biking to the metro station is beneficial to students' health. However, most such students are still escorted to and/or from school due to parental concerns about safety, comfort, reliability and convenience (Carver *et al.* 2013; Mandic *et al.* 2017). Since student metro users who live far from school rarely walk or cycle to school, they are more likely to be escorted by their parents if they are not using the metro on weekdays. For this reason, this article assumes that students who use the metro occasionally on weekdays are potential car passengers. Because metro use behaviour is a long term and dynamic process, it is very difficult capture through a traditional survey. Fortunately, the SC systems that are used for fare collection can provide detailed data, including the card type, boarding and alighting station and time spent on the metro for each user (Pelletier *et al.* 2011). Several studies have been conducted to investigate commuting behaviour using transit SCD. For instance, Long & Thill (2015) identified the residence, place of employment and commuting trip of SC holders using a 1-week period of SCD in conjunction with a one-day household travel survey. They assumed that transit commuters tend to take the metro regularly to and from similar sites and at similar times over long periods. Following that study, Long *et al.* (2016) further explored the spatio-temporal patterns of 4 types of extreme commuting behaviours using the same dataset. Ma *et al.* (2017) measured the spatial and temporal features of individual commuters, including residence, workplace, and departure time, to identify commuting patterns based on one month of SCD. Nonetheless, only a few studies have used SCD to analyse school commuting behaviours. Ordóñez Medina (2018) 1st proposed a method to identify study activity patterns based on the starting time and duration of students' commuting activities during one week. Using both SCD and a household travel survey, the author estimated the probability of activity types (home, study, other) and developed a discrete choice model for students. However, the study did not focus on the topic of school commuting behaviours by the metro from the perspective of escorting behaviour.

The aim of this study is to investigate this topic by: (1) identifying students who take the metro to school and examining their trips to/from school while considering the potential for escorting behaviour; (2) proposing a new method to measure potential escorting behaviour for school trips using SCD. The method can be used to identify the potential car trips of students who use metro transit occasionally beyond the use of a traditional survey. To achieve this objective, the factors influencing travel mode choice of schoolchildren must be understood. As noted above, the existing research studies on school commuting and escorting behaviour are based on survey data. For instance, Yarlagadda & Srinivasan (2008), Deka (2017), and Susilo & Liu (2016) found that parents' travel mode to work significantly influences the travel mode of their children. The use of a car for commuting increases the probability of escorting children. Li & Zhao (2015) found that students living in suburban areas travel longer distances, and cycling is more popular for students who live in urban areas. Zhang *et al.* (2017) demonstrated that students who travel a long-distance between home and school prefer motorized transport, and students living within their school's district have a lower probability of being driven to school. Although the data structure of SCs is dramatically different from that of traditional survey data, the same temporal and spatial characteristics of escorting behaviour can be adopted in our study. Therefore, a household travel survey is also used in this article.

2. Materials and methods

2.1. Study area context

Nanjing, the capital of Jiangsu Province (China), is used for this case study because (1) it is a typical large Chinese city that shares similar characteristics with other Chinese cities, and (2) parents in Nanjing attach great importance to their children's school education. Nanjing's population exceeds 8.2 million, its per capita gross domestic product was 127264 CNY (nearly 18927 US dollars) in 2016, and its total area is 6587 km^2 (Nanjing Statistics Bureau 2015). The city can be divided into 3 regions, namely, the urban, suburban, and exurban regions (Song *et al.* 2018). The urban area of the city can be characterized as mixed land use, with high densities of jobs and schools near rail transit facilities, while the exurban area is composed of single-function land use and has a low population density and poor rail transit facilities. By the end of 2016, there were 6 operating metro lines in Nanjing: Line 1, Line 2, Line 3, Line 10, Line S1, and Line S8. The total length of the lines is 225 km, and the total number of stations is 113, of which 16 stations are located in urban areas, 58 in suburban areas, and 39 in exurban areas, as shown in Figure 1a. We calculate the number of schools (including primary school, junior middle school and high school) and residences near the stations in the different regions. In addition, circles with a radius of 800 metres were adopted to identify schools and residences that are located near metro stations (Zhao, Deng 2013). Given this parameter, 58 schools and 424 residences are near metro stations in the urban area, 54 schools and 607 residences are near metro stations in the suburban area, 23 schools and 191 residences are near metro stations in the exurban area, as shown in Figure 1b.

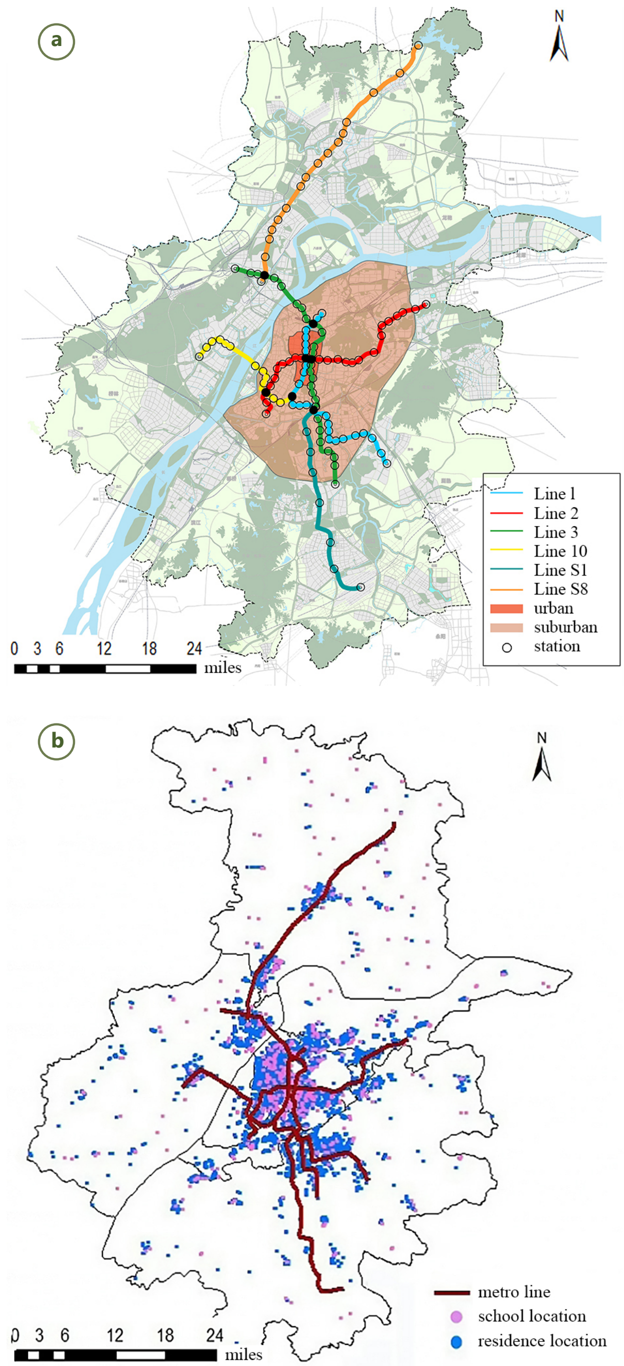

Figure 1. Study area of Nanjing:
(a) – regions and existing metro lines;
(b) – distribution of schools and residences

2.2. Data

Data from the SCs for the Nanjing public transit system can be used to obtain the card type, the station ID and time stamps for each card when a passenger enters or alights a metro station. Only students' cards were used in this study to achieve our research aim. Student SCs are available for students under 18 years old, including primary school students, junior high school students and high school students. Since the data were collected without passenger identification information because of privacy concerns, some identifiable characteristics of the respondents are unknown, such as students' ages and grades. The data used in this research were SCD for Nanjing rail transit from 1 to 31 October 2016; this period comprised fifteen specific weekdays from which samples of trips were extracted. The SCD set used in this article was a compilation of approximately 0.36 million transactions made by nearly 29 thousand SCs.

To identify school–housing locations and school commuting trips, the 2015 Nanjing resident travel survey was also used in this study. The survey included one-day trips by 12147 individuals in 31351 households of Nanjing, with an approximately 1% sampling rate. Respondents were asked to record their activity and travel information over 24 h. For each trip, the survey records the trip purpose and transport mode, start time/location, end time/location, and other important information, such as the latitude and longitude and the destination building type. In this study, only data on students studying in primary school, junior high school and high school and those who travelled to and from school on the survey day were extracted, which included 5915 students. Statistical analysis shows that 54.7% of the students were dropped off at school by their parents or other household members, and 20.6% of the students who were dropped off by car had a longer school commuting distance than other modes of transport. This result confirms the assumption that students who use metro transit occasionally on weekdays are potential car passengers.

2.3. Measures

2.3.1. Identifying and classifying school commuters

As stated in the Introduction, the research methods used to identify commuters and their commuting trips based on SCs have become mature; however, there are few studies on the identification of school commuting trips. Although Ordóñez Medina (2018) identified study activity patterns by the start time and duration of students' study activities during one week, the method used is limited to a travel survey and cannot reflect students who use the metro for one-way trips. Therefore, an identification method for students and their commuting trips considering escorts will be proposed in this article.

In this study, a metro school commuter is defined as a student who travels to and/or from school on weekdays using the metro. In most cities in China, including Nanjing, the school commutes of students going to school on weekdays are based on the school schedule shown in Figure 2. The vast majority of the students' trips to go to school on weekdays are subject to strict school discipline and parental management. Notably, some of the high school students in Nanjing are resident students. These students usually go to school on Sunday afternoon, stay at the school through the weekdays, and go home on Friday afternoon every week. Consequently, metro-commuting students can be identified based on the school schedule and the students' travel frequency. Nevertheless, there are still occasional non-commuting trips on weekdays by these identified students, and this limitation should be considered when evaluating the accuracy of the results. Note that we can assume the identified home station and school station as the student's home and school because according to the 2015 travel survey, over 90% of the students walked to the metro station. Each transaction contains the following fields:

- CID: the unique SC ID;
- De: the entry date (the date the SC holder entered the station);
- Te: the entry time (the time the SC holder entered the station);
- Dd: the departure date (the date the SC holder departed the station);
- Td: the departure time (the time the SC holder departed the station);
- SIDe: the metro station ID at entry;
- SIDd: the metro station ID at departure;
- Ct: card type (54 represents a student card);
- Long: longitude of the entry station;
- Lat: latitude of the entry station.

The metro system in Nanjing allows free transfers between 2 different lines, so each transaction record covers a complete trip by the metro. The critical step is to identify the stations where the school and home are located for each student based on some spatial and temporal rules, and then the school commuting trips can be identified from the trips between these 2 stations. Initially, the whole month's data were used. The following steps describe the identification process for commuting trips between home and school:

- *Step 1*: rebind all of the transactions and select records whose Ct is 54 (student card). Then, remove those transactions with the same SIDe and SIDd;
- *Step 2*: for each CID, calculate the frequency of the stations used, then find the station with the highest frequency and determine the number of stations with the highest frequency for each CID;
- *Step 3*: if the number of the station with the highest frequency is 1, then that station is a candidate for the student's home station or school station Si1; if the number is 2, then the 2 stations are the 2 candidates for the student's home station and school station Si1 and Si2; if the number is greater than 2, then merge the adjacent stations and find the station with the highest frequency as the home or school candidate Si1 or Si1 and Si2; if there are still more than 2 stations with the highest frequency, delete these travel records;
- *Step 4*: for those CIDs in which the number of stations with the highest frequency is 1 or greater than 2 in step 3, calculate the frequency of the station's use corresponding to the candidate station and then find the station with the highest frequency. If the number of the station with the highest frequency is 1, then the station is the other candidate for the student's home station or school station Si2; if the number of the highest frequency station is greater than or equal to 2, merge the adjacent stations and then identify the station with the highest frequency as the home or school candidate Si1.

Smart card data

CID	De	Te	Dd	Td	SIDe	SIDd	Ct
9965000xxxxx	2016/10/10	8:45:41	2016/10/10	9:01:38	25	5	54
9907780xxxxx	2016/10/10	8:40:59	2016/10/10	9:01:41	24	5	52
9965708xxxxx	2016/10/12	9:19:58	2016/10/12	9:30:59	10	14	53

School schedule

School type	To school	Out of school	To school	Out of school
Primary school	7:00–8:30	11:30–12:00	13:00–13:30	14:00–17:00
Junior high school	7:00–8:00	11:30–12:00	13:30–14:00	16:30–18:30
High school	6:30–7:30	11:30–12:00	13:30–14:00	17:30–21:30

Station location

SID	Long	Lat
25	118.77	32.04
24	118.76	32.04
10	118.78	32.05

Figure 2. SC transaction record and school schedule in Nanjing

If there are still greater than or equal to 2 stations with the highest frequency, delete these travel records;

- *Step 5*: transactions between the 2 candidate stations of each CID are divided into 4 categories on the basis of entry time: (1) AM, which refers to the morning before 9:00 am; (2) Noon1, between 11:30 am and 13:00 pm; (3) Noon2, from 13:00 pm to 14:00 pm; (4) PM, after 14:00 pm;
- *Step 6*: sort the transactions for each CID in chronological order. For the 1st and 3rd categories, define the 1st transaction's entry station as the home station and the other one as the school station. For the 2nd and 4th categories, define the last transaction's departure station as the home station and the other one as the school station. In this step, only students who travel from 9:00 am to 11:30 am are removed because it is difficult to determine whether they are commuting to school;
- *Step 7*: remove trips that started from school before 9:00 am and from home after 16:00 pm, which are assumed to be non-commuting trips;
- *Step 8*: delete the transactions for CIDs that commute between home and school by metro on only 2 days or less in 15 weekdays because it is difficult to determine if the school or home station has been identified by the above method, as the number of trips is too small and there is a lack of regularity.

The time rule is interpretable. The latest time to go to school (9:00 am and 14:00 pm) and the earliest time out of school (11:30 at Noon1 and 14:00 in the PM) in Nanjing's primary schools, junior high schools, and high schools are chosen as the temporal constraints, and students who use the metro before the latest start time or after the end time are probably school commuters. A small number of CIDs and their transactions for which it is difficult to identify the school and home station were deleted in order to obtain more accurate results. A flowchart is displayed in Figure 3.

Because the frequency of students using the metro for school commuting can reflect the situation of their escorts

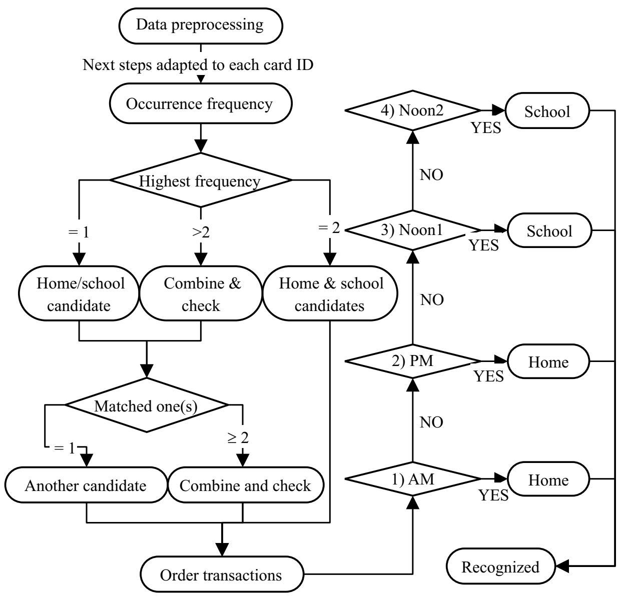

Figure 3. School commuter identification process

based on the assumption in the Introduction (Section 1), the lower the frequency, the more likely it is that the students are escorted by their parents. Although the frequency of students using the metro differs, they show similar travel behaviour in the same frequency groups. Thus, metro-commuting students are divided into different groups based on their usage frequency using *k*-means clustering in this article due to its good performance partitioning datasets into a number of clusters (Hartigan, Wong 1979).

2.3.2. Extracting escort-related travel characteristics

To study the escort behaviour of the different student groups, we need to extract their travel characteristics that are related to their escorts.

1st, school commuting behaviour is a round-trip process that includes trips to and from school. Thus, escorting involves picking up students from school, dropping off students at school and chauffeuring students both to and from the school. If a student's SCD only stores a trip record of a single direction (going to school or returning from school), we can assume that on the other trip, the student is escorted by his/her parents. Thus, this article defines the usage pattern of the metro for school commuting to reflect the escorting pattern that parents use to escort their children. 3 types are identified: HTSO refers to using the metro only when going to school from home, STHO refers to using the metro only when going home from school, and SHUTTLE means using the metro both to go to and return from school.

2nd, the synchronism of the departure time for parents and their children determines the possibility of escorting to a great extent (Mammen *et al.* 2014; He 2013; Liu *et al.* 2017). When the departure time for children's study activities is close to that of their parents' work activities, the parents are likely to escort their children on their commute. Departure time can be replaced with entry time at the home station and entry time at the school station because the home station and school station are assumed to be the home and school, respectively, in this article.

3rd, travel distance has a close relationship to parents' choice of escorting mode, as mentioned in the Introduction. As the distance between home and school increases, parents are more likely to use motorized modes of transport to escort their children. This can largely be expressed by the travel duration on the metro ($Tt = Td - Te$) due to its same or similar speed to escorting (Tt represents the duration of the metro ride, Td is the exit time for the SC holder, and Te is the entry time).

Additionally, the home location (Zhao *et al.* 2015) is also a key factor for parents' decision whether to escort their children. The transport mode for school commuting for students who live in the urban core area is significantly different from that for students outside the urban core area. Given the same travel distance, students who live outside the urban core area are more likely to be escorted by car. In reality, the school, as a major activity location for both students and parents, should also be taken into consideration. For this reason, it is necessary to explore the school–housing relationships between different student groups.

2.3.3. Multinomial logistic regression model

To generalize a logistic regression for the 3 student groups, multinomial logistic regression was performed in this study. We assume that a student i can be placed into frequency group k with maximum utility (McFadden *et al.* 1977) based on the metro usage pattern, entry time and school–housing relationship. Assume there are K frequency groups. The utility for student i who prefer the kth frequency group can be defined by the utility function U_{ik}, $k = 1, ..., K$, formulated as $U_{ik} = v_{ik} + \varepsilon_{ik}$. In this equation, v_{ik} represents the observable utility attribute, while ε_{ik} is an unobserved attribute. When ε_{ik} following a Gumbel distribution, the probability that student i belongs to group k can be expressed as follows:

$$p(i=k) = \frac{e^{(v_{ik})}}{\sum_{m \in K} e^{(v_{im})}}.$$

3. Analysis and results

3.1. Identifying and classifying school commuters

There are 72728 students in total, 28925 of whom were identified as commuters, accounting for 40%. After identifying commuting trips, the number of days each student used the metro and commuted by the metro within 15 continuous weekdays was calculated and called NOUD and NOCD, respectively. The distribution of the number of days that students used the metro and commuted by the metro during the 15 weekdays is shown in Figure 4 (student cards that showed commutes to school by the metro on less than 3 days in the 15 weekdays were deleted). The vast majority of students only used the metro occasionally on weekdays, with frequencies of less than 3 days/15 days. The frequency of most students who used the metro was distributed from 4 to 6 days and 13 days to 15 days, while a few students regularly but not fully used the subway, and their frequencies ranged 7 to 12 days. The regular pattern of trips commuting to school appears to entirely accord with the frequency distribution of the number of travel days for all students. This demonstrates the reliability of the above identification methods. No abnormal results were produced based on the identification processes described above.

To use the k-means algorithm, the number of clusters must be chosen beforehand. We created 2...4 categories based on the sample distribution. Comparing the clustering results for different groups, 3 types of student groups were finally determined. There are 11043 commuters whose NOCD is less than 6 and are identified as low-frequency, intermediate-frequency includes 8030 commuters with 6 to 12 NOCD, and the remaining 9852 commuters with more than 12 NOCD are considered high-frequency.

Among them, high-frequency users are the most representative type of commuters because they travel regularly and are the most loyal users of the metro. Additionally, intermediate-frequency users are considered to be potential regular metro users. Low-frequency users represent students who occasionally commute to school by metro in a more contingent and less regular manner, but the number of students in this cluster is relatively large.

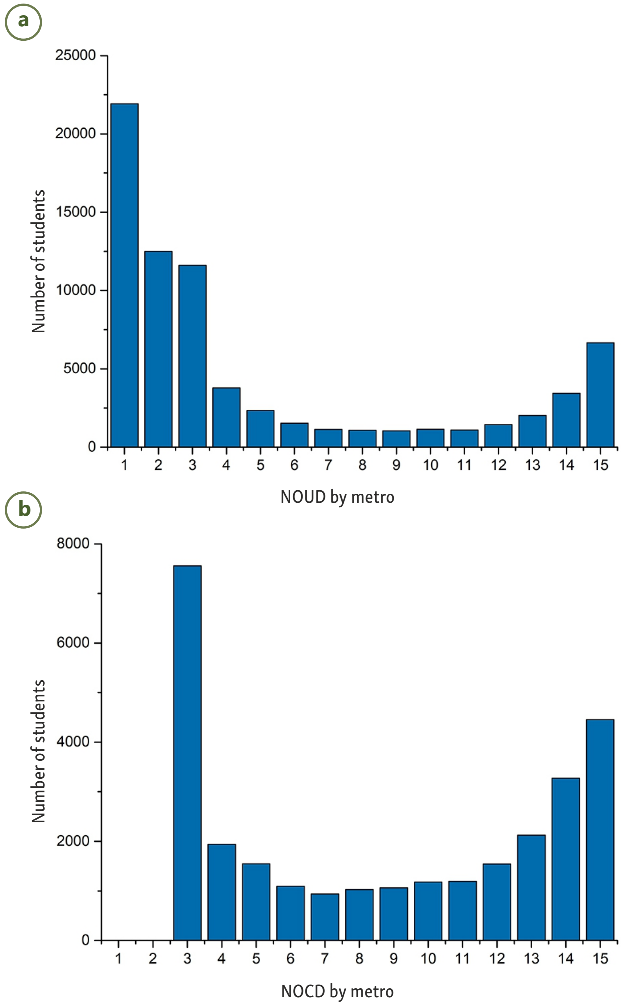

Figure 4. Distribution of the number of days students used the metro and commuted by the metro:
a – NOUD; b – NOCD

3.2. Extracting escort-related travel characteristics

3.2.1. Usage pattern characteristics

Figure 5 shows that there is a significant distinction between the different student groups. High-frequency users travel as regularly as normal work commuters do, and they are the most frequent users of the metro whose predominant pattern is SHUTTLE. However, the predominant pattern for intermediate-frequency and low-frequency users is STHO. This implies that these students are very likely to be escorted by their parents in the morning. This result can be confirmed with the 2015 Nanjing resident travel survey. In the travel survey, 52.7% of students were chauffeured to school by their parents, and they account for 92.6% of chauffeured students. However, the travel survey data do not reflect intermediate-frequency and low-frequency metro school commuters because almost all the records of students who use the metro for commuting show the SHUTTLE pattern in the survey data.

On the other hand, we find that the number of students in different metro usage patterns fluctuates slightly from day to day, especially for high-frequency and low-frequency users. For this reason, we use the ratio of the number of days for each pattern to the NOCD to obtain the long-term metro use pattern, which reflects the daily escorting situation. Statistically, 46% of students use the metro for STHO, followed by SHUTTLE (33.1%) and HTSO (15.4%). Note that because some boarders come home each Friday from their school, the number of low-frequency STHO students increases sharply on that day, as shown in Figure 5c.

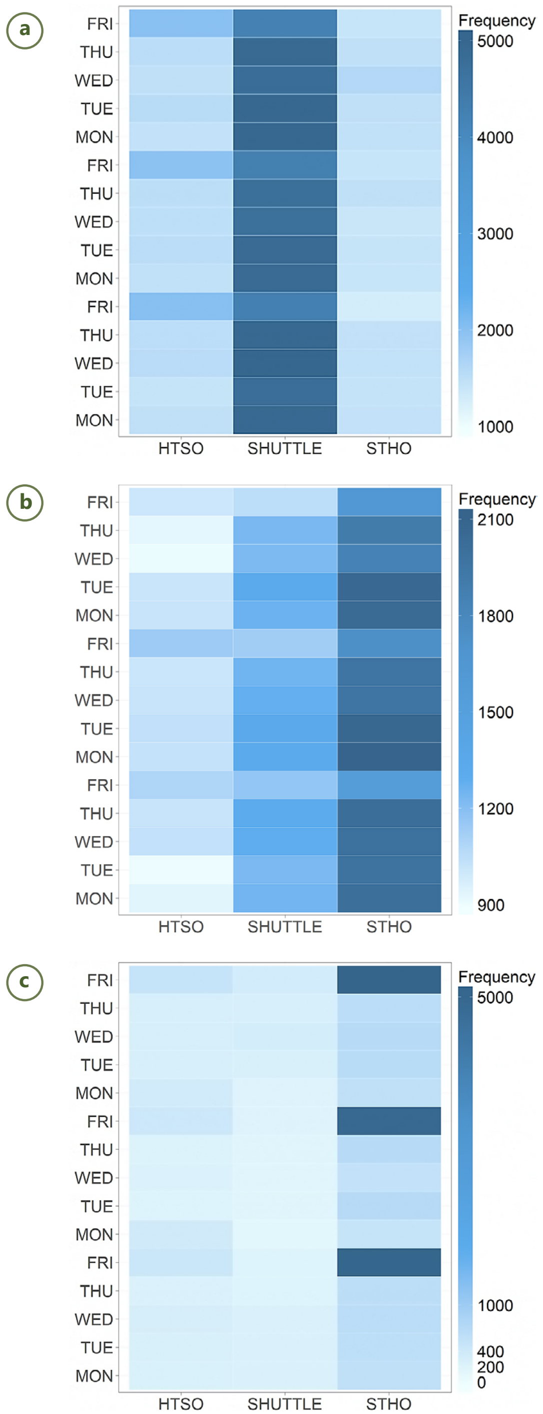

Figure 5. Usage patterns of the metro for school commuting by weekday: (a) – high-frequency; (b) – intermediate-frequency; (c) – low-frequency

3.2.2. Entry time characteristics

As Figure 6 shows, although there is only one morning peak of students travelling, the distribution of the entry times is different for the 3 student groups; 86.3% of the entry times for high-frequency users are between 6:10 am and 7:10 am, while 66.4% of the intermediate-frequency users and only 43.7% of the low-frequency users travel at the same time. The departure time of intermediate-frequency students and low-frequency students is later than that of high-frequency students and is mainly distributed between 6:20 am and 9:00 am. However, the entry time of the school station is more discrete than that of the home station. One possible reason for this result may be that the end of school hours differs for primary school, junior high school and high school.

3.2.3. Travel duration characteristics

Figure 7 shows that the difference between high-frequency and intermediate-frequency users is small, and approximately 85% of the travel times are less than 40 min (Figure 7a, Figure 7b). However, for low-frequency travellers, approximately 85% of the travel time is less than 57 min (Figure 7c), which is significantly longer than that for the other 2 types of commuters. This suggests that a long travel time to school also discourages students from using the metro. Thus, there should be an acceptable longest travel duration for students travelling to/from school by the metro. When the actual travel time by the metro exceeds the longest acceptable travel duration, the probability of students' metro use for school commuting decreases. Considering that the metro is the primary travel mode for high-frequency students commuting to/from school, the longest acceptable travel duration can be defined as the time that most (85%) of the high-frequency students are willing to tolerate. Additionally, public transport and private cars are the main travel modes for students who commute a long-distance to school. Thus, low-frequency students are more likely to be escorted by their parents in private cars when they do not choose to ride the metro to commute to school.

3.2.4. School–housing relationships

Based on the 3 identified regions, namely, urban, suburban and exurban, there are 9 combinations for the location of home and school in this article. For example, a student whose home is in an urban area and school is in a suburban area is represented by U–S. Statistically, students who live in the suburbs and go to school in urban and suburban areas account for the highest proportion (56.2%) of all metro-commuting students. One possible reason for this result is that most parents tend to live in the suburban or exurban areas due to the high price of housing in the urban area.

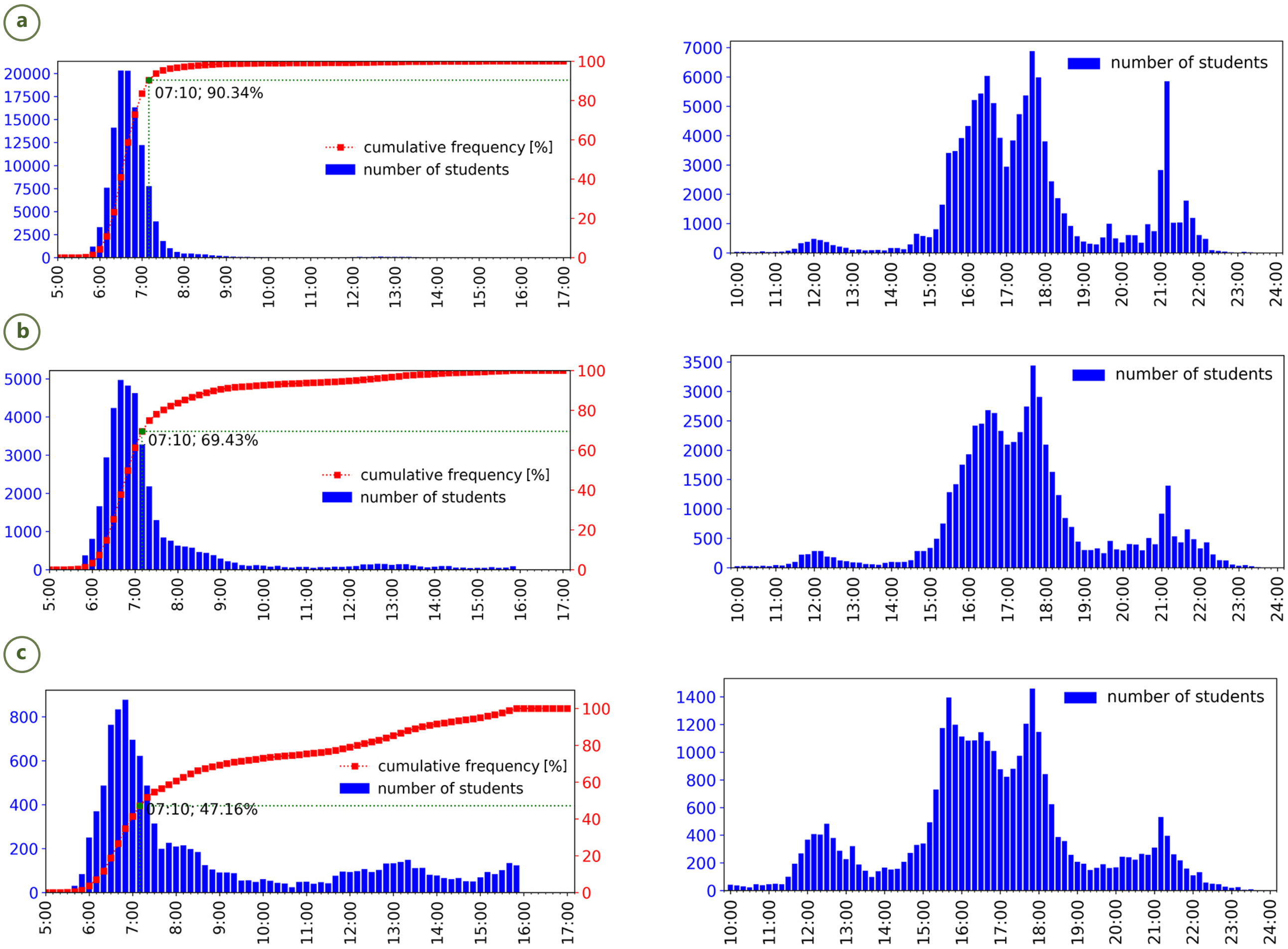

Figure 6. Departure time from home and school station:
(a) – high-frequency; (b) – intermediate-frequency; (c) – low-frequency

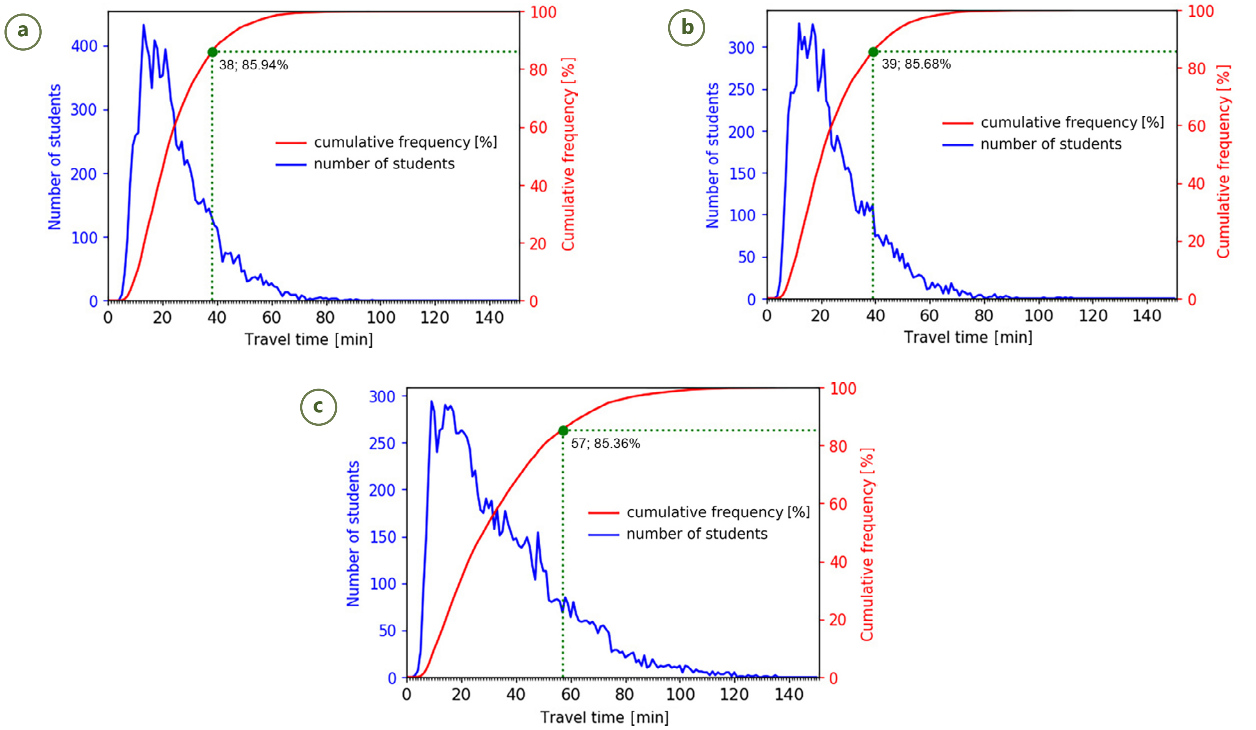

Figure 7. Travel time from home and school stations: (a) – high-frequency; (b) – intermediate-frequency; (c) – low-frequency

Another possible reason is that the urban and suburban areas have most of the educational resources, especially since the high-quality schools are mostly located in the urban area, even though that area is the smallest.

We also find that the location relationship between the home and school varies among the different groups. For both high-frequency and intermediate-frequency metro users, the home-school relationship is S–U (36.1% vs 28.2%), followed by S–S (26.6%. vs 21.6%). The number of other combinations is much lower. However, for low-frequency users, the most common relationship is U–S (21.2%) followed by S–U (17%) and E–S (12.4%). That is, those who have used the metro for a long time mainly live in suburban areas and go to school in urban or suburban areas, where the distance between metro stations is short. This result is consistent with Li and Zhao's (2015) findings that students living in the core area tend to go to school by public transport over a short distance. In addition, the proportion of E–E for low-frequency user is far higher than that for high-frequency users. These students are more likely to be escorted by their parents to and/or from school most of the time.

A summary and definitions of the independent variables that can affect the student groups are shown in Table 1.

Table 1. Explanation of variables

Variables	Definitions

Student groups	low-frequency = 1 (38.2%);
	intermediate-frequency = 2 (34.1%);
	high-frequency =3 (27.8%, reference group)

Usage patterns	OTHER = 1 (4.6%); SHUTTLE = 2 (33.1%); STHO = 3 (46.9%); HTSO = 4 (15.4%, reference group)
Entry time	continuous value
Travel duration	continuous value
School–housing relationships	U–U = 1 (7.2%); U–S = 2 (7.6%); U–E = 3 (1.2%); S–U = 4 (29.6%); S–S = 5 (26.6%); S–E = 6 (5.8%); E–U = 7 (4.2%); E–S = 8 (11.7%); E–E = 9 (6.1%, reference group)

3.3. Regression results

The estimation and evaluation results are reported in Table 2. As the *p*-value of the model is 0.00 (<0.05), the final model with selected independent variables is superior to the null model. The selected variables were significant at the 0.00 level, which indicates that usage pattern, entry time, travel duration and the school–housing relationship all have a significant impact on the grouping of students.

Usage pattern, as the main variable in this study, which reflects the commuting pattern among students using the metro, has a significant impact on the grouping results. Students who display the OTHER pattern are more likely to be intermediate-frequency users, with 1.27 times more than students who display the HTSO pattern. Students who display the STHO pattern are more likely to be low-frequency users than the other usage patterns. Students who display the SHUTTLE pattern are more likely to be high-frequency users rather than the other groups of users. In particular, the former students are scarcely likely to be low-frequency users. Entry time has a significant effect on the students' grouping between high-frequency and intermediate-frequency/low-frequency users. For every doubling of the entry time, the probability of students being low-frequency users increases 1.2 times. With regard

Table 2. Regression results

Mode	Intermediate-frequency			Low-frequency		
	Coefficient	Standard error	$p > z$	Coefficient	Standard error	$p > z$
Constant	-0.55^{***}	0.11	0.00	-1.03^{***}	0.11	0.00
OTHER	0.24^{***}	0.08	0.00	–0.13	0.09	0.13
SHUTTLE	-1.21^{***}	0.05	0.00	-2.11^{***}	0.06	0.00
STHO	-0.10^{*}	0.06	0.08	-0.44^{***}	0.06	0.00
Entry time	0.10^{***}	0.01	0.00	0.19^{***}	0.01	0.00
Travel duration	0.002^{*}	0.00	0.09	0.03^{***}	0.00	0.00
U–U	0.14	0.10	0.15	-0.73^{***}	0.10	0.00
U–S	-0.19^{*}	0.10	0.06	-0.93^{***}	0.10	0.00
U–E	–0.04	0.26	0.89	0.01	0.22	0.96
S–U	–0.12	0.08	0.15	-1.43^{***}	0.08	0.00
S–S	–0.12	0.08	0.14	-1.01^{***}	0.08	0.00
S–E	-0.23^{**}	0.11	0.04	-0.42^{***}	0.11	0.00
E–U	0.20^{*}	0.11	0.08	-1.09^{***}	0.12	0.00
E–S	0.07	0.10	0.49	-0.60^{***}	0.09	0.00
N = 28925; χ^2 = 12900.01; p = 0; log likelihood = –25061.26; McFadden pseudo R^2 = 0.21						

Notes: * – significance at 10%, ** – significance at 5%, *** – significance at 1%.

to travel duration, a longer travel duration contributes to a lower frequency of students using the metro for school commuting. In terms of the school–housing relationship, students who live and go to school in urban areas are less likely to be low-frequency users, at only 0.48 times the rate for students who live and go to school in exurban areas. Students who live in a suburban area and go to school in an urban area have only 0.24 times the probability of being low-frequency users compared to students who live and go to school in exurban areas. However, no statistically significant difference exists between S–U and E–E among the intermediate-frequency users. Similar results can be found for the other school–housing relationships.

4. Discussion

Students' metro use frequency for long-distance school commuting is important for understanding their potential escorting behaviour. This article focuses on different frequency of metro use groups of school commuting students using data from Nanjing. The results show that there are 11043 low-frequency students whose metro-commuting days are fewer than 6, intermediate-frequency users include 8030 students with 6 to 12 commuting days, and the remaining 9852 students with more than 12 commuting days are high-frequency metro users. Furthermore, the influencing factors of the student frequency groupings are identified using a multinomial logistic regression model. The results show that usage pattern, entry time, travel duration and the school–housing relationship are escort-related characteristics that have a significant impact on the grouping of students.

As expected, the metro use frequency for school commuting is closely related to the students' usage patterns. Students who display the SHUTTLE pattern are more likely to be high-frequency users rather than other types of users. Meanwhile, students who display the HTSO pattern are more likely to be low-frequency users than the other types of users. Intermediate-frequency users tend to use the metro based on the STHO pattern. One possible reason for these results may be that parents tend to drop off their children in the morning when they share the same departure time as their children, whereas they are less likely to pick up their children because they have different departure times. The entry time results further confirm this speculation. In terms of entry time, the later the departure time is, the more likely it is for students to be escorted to school. According to the results of the 2015 Nanjing resident travel survey, the commuting peak in Nanjing is centred at 7:00 am to 8:00 am and 8:00 am to 9:00 am. The results show that the departure time of those who use the metro is earlier than that of other commuters. However, the average departure time for parents who drop their children off on their commute is 7:18 am according to the survey data. This implies that some parents leave early to drop their children off at school because the departure time of most commuters is later. Since the departure time of the high-frequency students is 1 h earlier than that of their parents, it can be suggested that the departure time difference that parents find acceptable is approximately 1 h. Therefore, the implementation of staggered shifts of 1 h for employees and students may reduce the number of unnecessary escorting trips, which could help relieve the amount of traffic congestion during the AM peak time. This finding accords with the research of Zong *et al.* (2013).

Travel duration is a significant factor that influences students' metro use frequency for school commuting. A longer travel duration tends to contribute to a lower frequency of metro use for school commuting. In particular, as noted above, a student's longest acceptable travel duration is 38 min; thus, a school commuting time of more than 38 min is a main barrier to students' use of the metro. However, more than 85% of the students in Nanjing currently spend 45 min on the metro, which is 7 min longer than the longest acceptable travel duration. Therefore, reducing the duration of students' metro use would be conducive to increasing metro use for schooling. Furthermore, nearly all the high-frequency commuters live in suburban or urban areas and go to school in urban areas, while most of the low-frequency commuters need to travel across the region from the periphery to the centre. The results also reveal that a long school–housing distance significantly contributes to parents' escorting car trips. Therefore, there are 3 ways to shorten school commuting times. 1st, a more equitable distribution of educational resources in the city can be promoted by enacting relevant policies so that the catchment enrolment policy can truly play a role in providing convenience and travel guidance for students. As noted previously, most of the schools are located in areas within 800 m of subway stations, even though the urban area is the smallest. The suburban area and exurban area are much larger than the urban area, but the quantity of schools near subway stations is much less than in the urban area, especially the exurban area. Therefore, setting up new schools within a range of 800 m from subway stations could to some extent balance the city's educational resources and thereby reduce parents' escorting car trips. 2nd, in urban planning, the school–housing balance should be taken into account, especially in suburban and exurban areas, to reduce cross-regional or long-distance travel. Given that the actual travel time is 7 min longer than the longest acceptable travel duration, and the average speed of the Nanjing subway is 35 km/h, most students live 4 km farther than the ideal home-school distance. Finally, it is also important to continue to improve the accessibility of metro stations, and the last-mile issue can be resolved by means of, for example, bike sharing or minibuses, with well-designed traffic facilities for cyclists and pedestrians.

5. Conclusions

This study represents an effort to develop a method using SCD and a household travel survey to examine potential car-escorting trips for students who use metro transit beyond a traditional survey. Specifically, a school metro-commuter identification process was proposed by mining spatio-temporal travel regularities based on SCD during a continuous 3-week period. 3 frequency groups of school commuters were clustered based on their commuting days by the metro to identify the possibility of escorting behaviour. 4 possible factors influencing the clustering of the 3 frequency groups were extracted from SCD, including usage pattern of the metro, entry time, travel duration and the school–housing relationship. Furthermore, a multinomial logistic regression model was used to examine the significant factors that influence the grouping of students. Policy implications were generated from the results.

The contribution of this article lies in 3 aspects. 1st, this study solves the problem of identifying metro-commuting students while considering students who use the metro for only one trip based on SCD and thus provides a method for the use of metro big data and travel survey data. This approach can identify the potential escorting behaviour behind students' school commuting by metro, which is difficult to obtain though a traditional household travel survey. 2nd, 4 escort-related characteristics that were extracted from 3 weeks of SCD can be used to help identify frequency of metro use groups of respondents from one-day household survey data. 3rd, recommendations for planners and policymakers for school travel demand management are proposed to reduce parents' car usage for long-distance escorting. Overall, this study fills a gap in the literature on school travel by using SCD and a household travel survey and proposes a new research perspective for escorting behaviour.

Future research could extend this study to analyse more profound factors that may affect students' school commuting, such as the socioeconomic attributes of card holders and their parents, by combining a questionnaire and SCD. Since the data were collected without passenger identification information due to privacy concerns, the ages of the card holders were unknown. We also could not classify students according to their grades, and as a result, targeted suggestions based on school grades could not be proposed in this article. Therefore, this should be considered as an aspect of future work.

Acknowledgements

We thank *Dr. Zhuangbin Shi* for his helpful revision on final version of article before submission.

The authors also express their gratitude to the *Nanjing Metro Group Co Ltd* for providing transaction data.

Funding

The work was supported by:
- the Project of National Natural Science Foundation of China under Grant No 52202381;
- the Fundamental Research Project of Yunnan Province (China) under Grant No 202201AU070109.
- the Scientific Research Foundation of Graduated School of Southeast University under Grant No YBJJ1842;
- the China Scholarship Council under Grant No 201806090205.

Author contributions

Yanjie Ji and *Yang Liu* conceived and designed the study.

Yanjie Ji and *Yang Liu* were responsible for data collection and analysis.

Yang Liu and *Xinwei Ma* were responsible for data interpretation.

Yang Liu and *Qiyang Liu* wrote the 1st draft of the article.

Yang Liu revised the article.

Disclosure statement

This article has not been published and is not under consideration for publication elsewhere.

We have no conflicts of interest to disclose.

References

Carver, A.; Timperio, A.; Crawford, D. 2013. Parental chauffeurs: what drives their transport choice?, *Journal of Transport Geography* 26: 72–77. https://doi.org/10.1016/j.jtrangeo.2012.08.017

Deka, D. 2017. Impacts of standardizing school start time on children and household workers – an examination with NHTS data, *Journal of Transport Geography* 59: 40–48. https://doi.org/10.1016/j.jtrangeo.2017.01.005

Easton, S.; Ferrari, E. 2015. Children's travel to school – the interaction of individual, neighbourhood and school factors, *Transport Policy* 44: 9–18. https://doi.org/10.1016/j.tranpol.2015.05.023

Hartigan, J. A.; Wong, M. A. 1979. Algorithm AS 136: a *k*-means clustering algorithm, *Journal of the Royal Statistical Society*. *Series C* (*Applied Statistics*) 28(1): 100–108. https://doi.org/10.2307/2346830

He, S. Y. 2013. Will you escort your child to school? The effect of spatial and temporal constraints of parental employment, *Applied* Geography 42: 116–123. https://doi.org/10.1016/j.apgeog.2013.05.003

He, S. Y.; Giuliano, G. 2018. School choice: understanding the trade-off between travel distance and school quality, *Transportation* 45(5): 1475–1498. https://doi.org/10.1007/s11116-017-9773-3

Li, S.; Zhao, P. 2015. The determinants of commuting mode choice among school children in Beijing, *Journal of Transport Geography* 46: 112–121. https://doi.org/10.1016/j.jtrangeo.2015.06.010

Lingaitis, L. P.; Bureika, G.; Vaičiūnas, G. 2003. The study of schoolchildren transportation efficiency in Lithuania, *Transport* 18(2): 66–71. https://doi.org/10.3846/16483840.2003.10414068

Liu, Y.; Ji, Y.; Liu, Q.; He, M.; Ma, X. 2017. Investigating electric bicycles as a travel mode choice for escorting children to school: a case study in Kunming, China, *Transportation Research Record*: *Journal of the Transportation Research Board* 2634: 8–16. https://doi.org/10.3141/2634-02

Long, Y.; Liu. X.; Zhou, J.; Chai, Y. 2016. Early birds, night owls, and tireless/recurring itinerants: an exploratory analysis of extreme transit behaviors in Beijing, China, *Habitat International* 57: 223–232. https://doi.org/10.1016/j.habitatint.2016.08.004

Long, Y.; Thill, J.-C. 2015. Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing, *Computers*, *Environment and Urban Systems* 53: 19–35. https://doi.org/10.1016/j.compenvurbsys.2015.02.005

Lu, M.; Sun, C.; Zheng, S. 2017. Congestion and pollution consequences of driving-to-school trips: a case study in Beijing, *Transportation Research Part D*: *Transport and Environment* 50: 280–291. https://doi.org/10.1016/j.trd.2016.10.023

Ma, X.; Liu, C.; Wen, H.; Wang, Y.; Wu, Y.-J. 2017. Understanding commuting patterns using transit smart card data, *Journal of Transport Geography* 58: 135–145. https://doi.org/10.1016/j.jtrangeo.2016.12.001

Mammen, G.; Stone, M. R.; Buliung, R.; Faulkner, G. 2014. School travel planning in Canada: identifying child, family, and school-level characteristics associated with travel mode shift from driving to active school travel, *Journal of Transport* & *Health* 1(4): 288–294. https://doi.org/10.1016/j.jth.2014.09.004

Mandic, S.; Sandretto, S.; García Bengoechea, E.; Hopkins, D.; Moore, A.; Rodda, J.; Wilson, G. 2017. Enrolling in the closest school or not? Implications of school choice decisions for active transport to school, *Journal of Transport* & *Health* 6: 347–357. https://doi.org/10.1016/j.jth.2017.05.006

McDonald, N. C. 2008a. Children's mode choice for the school trip: the role of distance and school location in walking to school, *Transportation* 35(1): 23–35. https://doi.org/10.1007/s11116-007-9135-7

McDonald, N. C. 2008b. Household interactions and children's school travel: the effect of parental work patterns on walking and biking to school, *Journal of Transport Geography* 16(5): 324–331. https://doi.org/10.1016/j.jtrangeo.2008.01.002

McFadden, D.; Train, K.; Tye, W. B. 1977. An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model, *Transportation Research Record* 637: 39–46. Available from Internet: http://onlinepubs.trb.org/Onlinepubs/trr/1977/637/637-007.pdf

Mitra, R.; Buliung, R. N.; Roorda, M. J. 2010. Built environment and school travel mode choice in Toronto, Canada, *Transportation Research Record*: *Journal of the Transportation Research Board* 2156: 150–159. https://doi.org/10.3141/2156-17

Nanjing Statistics Bureau. 2015. *2015 Nanjing National Economic and Social Development Statistics Bulletin*. Nanjing Statistics Bureau, Nanjing, China. (in Chinese).

Ordóñez Medina, S. A. 2018. Inferring weekly primary activity patterns using public transport smart card data and a household travel survey, *Travel Behaviour and Society* 12: 93–101. https://doi.org/10.1016/j.tbs.2016.11.005

Pelletier, M.-P.; Trépanier, M.; Morency, C. 2011. Smart card data use in public transit: a literature review, *Transportation Research Part C*: *Emerging Technologies* 19(4): 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Song, Y.; Fan, Y.; Li, X.; Ji, Y. 2018. Multidimensional visualization of transit smartcard data using space–time plots and data cubes, *Transportation* 45(2): 311–333. https://doi.org/10.1007/s11116-017-9790-2

Susilo, Y. O.; Liu, C. 2016. The influence of parents' travel patterns, perceptions and residential self-selectivity to their children travel mode shares, *Transportation* 43(2): 357–378. https://doi.org/10.1007/s11116-015-9579-0

Wen, L. M.; Fry, D.; Rissel, C.; Dirkis, H.; Balafas, A.; Merom, D. 2008. Factors associated with children being driven to school: implications for walk to school programs, *Health Education Research* 23(2): 325–334. https://doi.org/10.1093/her/cym043

Wilson, E. J.; Marshall, J.; Wilson, R.; Krizek, K. J. 2010. By foot, bus or car: children's school travel and school choice policy, *Environment and Planning A*: *Economy and Space* 42(9): 2168–2185. https://doi.org/10.1068/a435

Yarlagadda, A. K.; Srinivasan, S. 2008. Modeling children's school travel mode and parental escort decisions, *Transportation* 35(2): 201–218. https://doi.org/10.1007/s11116-007-9144-6

Zhang, R.; Yao, E.; Liu, Z. 2017. School travel mode choice in Beijing, China, *Journal of Transport Geography* 62: 98–110. https://doi.org/10.1016/j.jtrangeo.2017.06.001

Zhao, J.; Deng, W. 2013. Relationship of walk access distance to rapid rail transit stations with personal characteristics and station context, *Journal of Urban Planning and Development* 139(4): 311–321. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000155

Zong, F.; Juan, Z.; Jia, H. 2013. Examination of staggered shifts impacts on travel behavior: a case study of Beijing, China, *Transport* 28(2): 175–185. https://doi.org/10.3846/16484142.2013.803263