Transportation Research Part F 15 (2012) 38–51

Contents lists available at SciVerse ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

# Measuring rail passenger crowding: Scale development and psychometric properties

Nor Diana Mohd Mahudin *, Tom Cox, Amanda Griffiths

*Institute of Work, Health and Organisations, University of Nottingham, International House, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, United Kingdom*

ARTICLE INFO

*Article history:*
Received 12 November 2010
Received in revised form 23 November 2011
Accepted 28 November 2011

*Keywords:*
Crowding
Rail passenger
Stress
Scale development
Structural equation modelling

ABSTRACT

Research on rail passenger crowding often tacitly subscribes to a measurement of crowding based on density (i.e. physical conditions involving space limitations) and rarely considers the possible role psychological factors may play in measuring this construct. This paper describes the development of an instrument that captures the dimensionality of rail passenger crowding and its relationship to the experience of stress and feelings of exhaustion. The proposed instrument is a 20-item self-rating questionnaire consisting of three sub-scales designed to assess subjective crowding experiences among rail users ($n = 525$). Findings from the factor analyses generally support the hypothesised three-factor structure of the measurement model (evaluation of the psychosocial aspects of the crowded situation, evaluation of the ambient environment of the crowded situation, and affective reactions to the crowded situation). All sub-scales demonstrate excellent internal consistency and construct validity as well as good convergent and discriminant validity values. The instrument was further tested using structural equation modelling to examine the impact of crowding on commuters' stress and feelings of exhaustion. With the addition of the "passenger density" variable as an indicator of objective measurement of crowding operating in tandem with the crowding sub-scales, the results reveal that: (1) commuters' evaluations of the psychosocial aspects of the crowded situation and of its ambient environment, alongside their rating of passenger density, significantly predict affective reactions to the crowded situation; (2) these affective reactions, in turn, significantly predict stress and feelings of exhaustion; and (3) evaluations of the psychosocial aspects of the crowded situation and of its ambient environment as well as passenger density do not directly predict stress and feelings of exhaustion. The link between rail passenger crowding and the negative outcomes therefore does not appear as a simple, direct relationship, but is mediated by affective feelings of crowdedness. Overall, these results provide satisfactory psychometric properties for the proposed instrument and support its use as an assessment tool for measuring crowding experience in the rail setting.



## 1. Introduction

The impact of crowding on commuting experience is a major issue for public transport users, and more so for rail passengers. However, studies concerning this issue represent only a small part of the studies of commuting stress (e.g. Thomas, Rhind, & Robinson, 2006; Cassidy, 1992). Crowding has often been highlighted as one of the factors that can lead to the

* Corresponding author. Tel.: +44 0 115 846 7523; fax: +44 0 115 846 6625.
*E-mail addresses:* lwxndm@nottingham.ac.uk (N.D. Mohd Mahudin), tom.cox@nottingham.ac.uk (T. Cox), amanda.griffiths@nottingham.ac.uk (A. Griffiths).

1369-8478/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trf.2011.11.006

experience of strain in passengers, but which to date has been unsatisfactorily explained. One reason for the comparatively few studies addressing this issue is possibly related to the lack of a precise definition of crowding for use in the rail transport context, which in turn predictably leads to difficulties in measurement (Cox, Houdmont, & Griffiths, 2006; Turner, Corbett, O'Hara, & White, 2004). An exhaustive measure whereby the components of crowding that are accurately assessed from the passengers' viewpoint is therefore desired.

The term ''crowding'' has proven to be challenging because it has multiple meanings and has been inconsistently defined. In the general literature, it is typically referred to as the ratio of people to space, conceived in terms of density (Freedman, 1975; Fischer, Baldassare, & Ofshe, 1975). However, researchers such as Stokols (1972), argue that defining crowding based solely on density is somewhat inaccurate because crowding also has subjective elements to it. Taking up Stokol's argument, most researchers have now made a conceptual distinction between density and crowding (Cox et al., 2006; Baum & Paulus, 1987). The former, which is defined as a physical condition of having limited space available for the given number of individuals present, is considered as a necessary but not sufficient condition for crowding, whereas the latter is defined as a subjective, psychological state in which one's expectation of space exceeds the available supply (Stokols, 1972).

The conceptual definition of crowding remains a debatable issue in the rail literature, with the currently available studies revealing inconsistent use of crowding terminology and measurement by both the industry and the authorities (Turner, Corbett, O'Hara, & White, 2004). Up to this point, the literature indicates that rail crowding is defined most often as a quantitative and objectively verifiable experience, which may be based on measurements of passenger density and train capacity (Turner, Corbett, O'Hara, & White, 2004). Some examples of these measurements include seating and standing capacity (e.g. Dodgson, Kelso, Van der Veer, Skene, & Paredes, 2002), passenger loading based on levels of service (e.g. Lam, Cheung, & Lam, 1999; Lam, Cheung, & Poon, 1999) or passengers in excess of capacity (e.g. UK Department for Transport, 2009), or different percentages of train capacity such as degree of crowding (e.g. Lam, Cheung, & Lam, 1999; Lam, Cheung, & Poon, 1999). Previous attempts at assessing crowding experience on trains have also been primarily focused on the measurement of passenger density. For example, studies by Evans and Wener (2007), Buckley and O'Regan (2004), the UK Rail Safety and Standards Board (2004), Wener and Evans (2004), Singer, Lundberg, and Frankenhauser (1978), and Lundberg (1976) measured crowding by means of self-rating questionnaires in which participants were asked to rate levels of crowd density inside the trains.

Although density seems to have been the standard method for defining and measuring crowding in the rail industry, several researchers argue that this definition is somewhat limited in that it does not take into consideration factors such as an individual's subjective perception of crowdedness, which may account for variability in outcomes among rail users (Cox et al., 2006; Turner et al., 2004). Furthermore, treating density and crowding as interchangeable terms may potentially lead to confusion in the definition of both constructs and difficulty in their measurements. Turner et al. (2004) therefore conclude that the definition of crowding in rail settings should encompass two dimensions, focusing on (1) density and the available space, which represent the objective components, and (2) perceptions of both the available space in the physical condition and the number of people present, which characterise the subjective elements. A comprehensive measure based on both subjective and objective elements of crowding is thus called for from this point of view.

To date, a standardised instrument has yet to be developed that assesses the different psychological components of crowding. At present, our knowledge is derived from a relatively small number of studies which tend to be conducted in the residential or retail setting. For example, the scale used in Kaya and Weber's study (2003) is made up of six items with a seven-point Likert scale measuring crowding experience in a residence hall environment and ten items with a semantic differential response scale that assesses the perception of hall room. Meanwhile, the scales used in studies by Evans, Lepore, and Mata-Allen (2000) and Fuller, Edwards, Sermsri, and Vorakitphokatorn (1993) consist of items that ask about household crowding in terms of the level of crowding experienced and restricted movement. Crowding scales developed in retail studies such as those by Harrell, Hutt, and Anderson (1980) and Machleit, Kellaris, and Eroglu (1994) also more or less consist of similar dimensions of confined or closed-in feelings and restricted movement.

Only a few attempts have been made to distinguish the measurement of physical density from that of the subjective evaluation of crowding. For instance, Kalb and Keating (1981), who constructed a scale consisting of 28 items with a ten-point bipolar semantic differential response scale that measures how crowded people feel, how crowded people rate a setting, and how people rate the crowding ambience, found that these three components can be attributed to individuals' experience of crowding. In particular, they discovered that self-ratings of crowding experienced and description ratings of crowdedness are conceptually distinct, since the feeling of crowdedness component is loaded with perceived density, constraint, distraction, and stress. On the other hand, the description rating component is loaded only with perceived density and general negative affects. Relatively similar components are also identified to be distinctive in Mohd Mahudin's (2003) study. According to this study, crowding can be described and measured in terms of eight components: physical, social, personal, psychological, crowd characteristics, crowd dynamics, crowd behaviour, and location. Four of the components, i.e. physical, social, personal, and psychological, relate to the feelings of crowdedness, whereas crowd characteristics, crowd dynamics, crowd behaviour, and location describe where crowding takes place.

While these scales have previously been identified as promising for assessing individuals' crowding experience, they had neither been verified nor tested for use in the rail passenger context. In addition, with the exception of scales by Machleit et al. (1994), Harrell et al. (1980), and Kalb and Keating (1981), a thorough analysis of the psychometric properties of these scales is neither conducted nor reported. Furthermore, measuring passengers' crowding experiences based solely on density

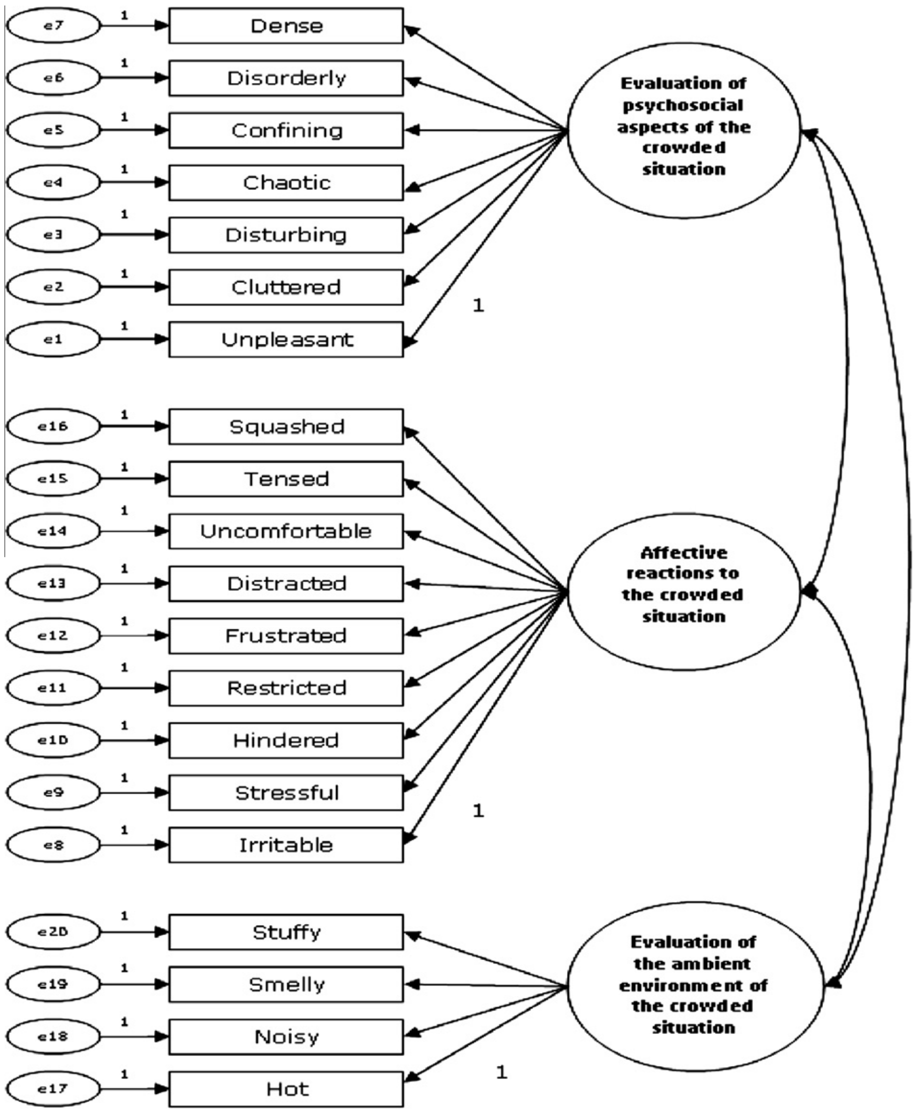


**Fig. 1.** The hypothesised measurement model of rail passenger crowding.

remains limited, particularly when examining individual, subjective perceptions of crowdedness, and fails to capture many of the outcomes that might occur as a result of crowding (Cox et al., 2006). In essence, the scales developed so far are relatively inadequate in tapping the overall dimensions of crowding experience. Consequently, there is a need to develop an accurate and reliable scale that can capture the various psychological components of crowding for use in the rail passenger context and possibly in other transportation settings as well.

The objective of this study is to develop and test an instrument designed to tap the different psychological components of crowding. Developing an instrument to measure this construct serves the threefold purpose of bridging theory, research, and practice. First, the development of such an instrument could clarify and make more concrete the meaning of rail passenger crowding, thus providing a specific operational definition of the construct. Second, a reliable and valid instrument that has good psychometric properties could lend itself to future research in the field. Finally, an instrument that adequately reflects the experience of crowding among rail users could be used as an assessment tool to aid train operating companies as well as relevant authorities in addressing the issue of passenger crowding.

The measurement and structural models of the proposed instrument are elaborated in Section 2, which is followed by a section on the methods and procedures undertaken in this study. The model estimation results are then reported in Section 4. Finally, Section 5 offers a discussion of the possible uses and implications of the instrument as well as providing a summary and conclusions for the study.

## 2. Modelling frameworks

### *2.1. Hypothesised measurement model*

A measurement model was first specified to evaluate the adequacy of the proposed model in explaining the underlying observed data. It was hypothesised that the experience of rail passenger crowding is defined by three different psychological components or latent factors, i.e. (1) evaluation of the psychosocial aspects of the crowded situation; (2) evaluation of the ambient environment of the crowded situation; and (3) affective reactions to the crowded situation; all measured by 20 indicator variables. The measurement model also assumes that the three latent factors are intercorrelated, that each item loads on one and only one factor, and that the measurement errors are not correlated with one another. This is depicted in Fig. 1.

### *2.2. Hypothesised structural model*

In a systematic review of 41 studies conducted by Mohd Mahudin, Cox, and Griffiths (2010), sufficient evidence was found to conclude that rail passenger crowding has a significant and consistent negative impact on passengers in terms of psychological stress or emotional distress. To evaluate the use of the proposed instrument in practice, a structural model accounting for the effects of rail passenger crowding on stress and feelings of exhaustion was tested (see Fig. 2).

It should be noted that although the present study focused on the roles that the different psychological components of crowding play in influencing negative outcomes experienced by passengers, the importance of the density variable could not be ignored. In the realm of rail studies, several researchers have differentiated between passenger density and perceived crowdedness or crowding (Cox et al., 2006; Turner et al., 2004). Additionally, Evans and Wener (2007) found that the immediate, close presence of other passengers in a train car is more salient for the experience of crowding than the overall train density. From the work of Kalb and Keating (1981), it can be suggested that further distinctions are made between the crowding experience, the essentially psychosocial characteristics of the crowded situation, and those of its ambient environment. It is argued that all three components of crowding, together with passenger density, may interplay and interact to exert their effects on individuals. Understanding the interplay between these variables could offer insight into the nature of rail passenger crowding as studies have shown that high-density situations may affect both perceived crowding (Cox et al., 2006) and induce affective feelings of crowdedness (Bell, Green, Fisher, & Baum, 2001; Baum & Paulus, 1987). Following this line of reasoning, the passenger density variable was included in the structural model analysis. As presented in Fig. 2, evaluations of the psychosocial aspects of the crowded situation and of its ambient environment, alongside passenger density, are hypothesised to influence affective reactions to the crowded situation (H1–H3). These affective reactions to crowding are then expected to influence the experience of stress (H4) and feelings of exhaustion (H5).

## 3. Method

### *3.1. Initial item development and pilot study*

The initial list of possible items for measuring the psychological components of crowding was generated by surveying existing scales and reviewing the literature. A single scale which comprised of nine items was then developed and piloted

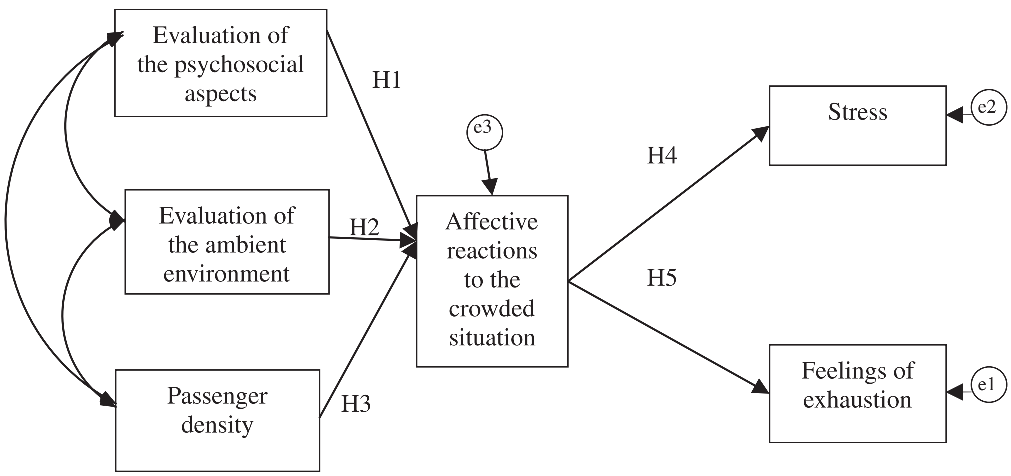


**Fig. 2.** The hypothesised structural model representing the relationships among the three latent factors of rail passenger crowding, passenger density, experience of stress, and feelings of exhaustion.

on 25 individuals who commuted to work regularly and frequently via the light rail transit or commuter rail services. The respondents were asked to rate their agreement with statements such as "*The train I commute on is usually crowded*" and "*There is not enough space on the train for commuters*", using a six-point Likert scale ranging from "Strongly disagree" (scored 1) to "Strongly agree" (scored 6), with higher total scores indicating a greater experience of crowding. Besides filling in the questionnaire, the respondents were asked to identify the items that they viewed as particularly important by indicating whether or not each item is "essential" to the overall study. These responses were measured using Lawshe's (1975) content validity ratio (CVR) formula in order to explore how each item is judged to be reflective of, or theoretically linked to, the crowding experience construct. The CVR value ranges from −1 to +1, with values closer to +1 indicating that the respondents agreed that the item is "essential" and, therefore, valid. In addition to this protocol, the respondents completed a supplemental section that asked them to comment on the wording and ordering of the items as well as on the overall questionnaire format and layout.

### 3.2. Item refinement

Only items that exceed the minimum recommended level of Cronbach's coefficient alpha of .70 (Nunnally & Bernstein, 1994) and have a CVR value higher than .49 were retained (following Lawshe's (1975) recommendation on the minimum CVR value for 15 panel members). Two specific points of interest were highlighted in the qualitative comments from the respondents. These comprise the importance of categorising the items according to the hypothesised dimensions and the need for a better and clearer response scale format for all items. In light of these suggestions, three decisions were made when refining the proposed instrument. First, it was decided to revise and group the items according to the three constructs being tapped. Second, the response scale format was changed from the strongly disagree-strongly agree format to a construct-specific response scale format. The latter format, which involves phrasing the response in terms of the construct being measured, was chosen because it has been proven to have greater accuracy, reliability, and stability compared to the disagree-agree format (Saris, Revilla, Krosnick, & Shaeffer, 2010; Ping, 2005). Finally, the three groups of crowding items were treated as separate scales in their own right. The basis for this decision stems from Kalb and Keating's (1981) and Mohd Mahudin's (2003) studies discussed earlier in Section 1.

The refined instrument used in the main study therefore consisted of three separate scales with a five-point construct-specific response scale format.[1] The first subscale was named "*Evaluation of the psychosocial aspects of the crowded situation*" with seven items, while the second subscale consisted of nine items and was named "*Affective reactions to the crowded situation*". Meanwhile, the third subscale comprised four crowding atmosphere items and was called "*Evaluation of the ambient environment of the crowded situation*". For each subscale, items were summed, with higher scores indicating a more negative experience of crowding.

### 3.3. Main study

#### 3.3.1. Participants

Data for the main study were collected by a questionnaire survey from a sample of 525 commuters on four rail lines serving Kuala Lumpur, Malaysia. These rail lines include two light rail transit (Kelana Jaya and Ampang) and two commuter rail (KTM Komuter Rawang-Seremban and KTM Komuter Sentul-Klang) services. Similar recruitment criteria adopted in the pilot study were used to screen the respondents, i.e. only people who commute to work frequently using the light rail transit or commuter rail services for at least 3 days per week and have been on the same route for at least 6 months were eligible to participate. The majority of the respondents (51.6%) were Kelana Jaya line users, followed by the Ampang line (17.9%), KTM Komuter Rawang-Seremban (9.9%), and KTM Komuter Sentul-Klang (3.8%). The remaining 16.8% of the sample was constituted of commuters who normally changed between trains or to other transportation to complete their journeys to work. The mean age of the respondents was 27.7 years (SD = 7.3), with more than half of the respondents (407 or 77.5%) reporting having at least a diploma, undergraduate or postgraduate degree. Most of the respondents were females (63.2%) and had been using the rail services for 1–4 years (46.9%), which indicates that the obtained sample resembles the population of rail users in Malaysia (Yusoff, 2008).

#### 3.3.2. Measures

In addition to the new rail passenger crowding instrument, the questionnaire also incorporated three other measures. First, as mentioned in Section 2.2, the passenger density variable was added to the structural model analysis as an objective measure of crowding. This variable was assessed using a pictorial device developed from that used by the UK Rail Safety and Standards Board (2004). Respondents were presented with a scale made up of four pictorial representations of commuters of increasing passenger density. They were asked to use this scale to rate the overall crowd density in the trains. Meanwhile, the outcome measures were assessed by the stress subscale of the Stress and Arousal Checklist (SACL: Gotts & Cox, 1988) and the worn-out subscale of the General Well-Being Questionnaire (GWBQ: Cox & Gotts, 1987). These scales were selected because they have been shown to be reliable and valid measures of experience of stress and feelings of exhaustion and had

[1] A copy of the complete *scale is available upon request* from the first *author*.

been used in previous studies of rail passenger experience (e.g. Jain, Cox, & Houdmont, 2006). High scores indicate increased psychological stress and greater feelings of exhaustion. In this investigation, the Cronbach's alpha scale reliabilities were $\alpha = .81$ for the stress subscale and $\alpha = .87$ for the worn-out subscale.

#### 3.3.3. Data analysis

The factor structure and internal consistency of the proposed instrument was tested using exploratory factor analysis and Cronbach's alpha coefficients. The hypothesised three-factor model was then examined to ensure that all variables in the model accurately corresponded to their intended latent factors. This generally involves assessing the consistency and validity of the constructs and indicator items. In particular, confirmatory factor analysis was carried out via AMOS 16.0 to assess the indicator reliability and construct validity as well as the convergent and discriminant validity of the crowding measures, and to test whether the experience of crowding can be explained by the three latent factors proposed. Since AMOS 16.0 can only compute standardised regression weights, computations for construct and discriminant validity such as composite reliability and average variance extracted were calculated separately in a spreadsheet using Hatcher's (1994) equations.

The adequacy or the goodness-of-fit of the hypothesised measurement model was computed using five fit indices: the chi-square goodness-of-fit statistic; the standardised root mean residual (SRMR); the root-mean-square error of approximation (RMSEA); the comparative fit index (CFI); and the parsimonious normed fit index (PNFI). Additionally, the parsimony of the model was assessed using Akaike's information criterion (AIC: Akaike, 1987). Altogether, these indices were selected over other indices because they have been found to be the most insensitive to sample size, model misspecification, and parameter estimates (Hooper, Coughlan, & Mullen, 2008).

The hypothesised structural model was subjected to path analysis testing using structural equation modelling in order to investigate the predictive validity of the crowding measures and passenger density on the experience of stress and feelings of exhaustion. This technique was chosen because it has the capability of assessing complex relationships between multivariate data in a single, systematic, and comprehensive analysis (Gefen, Straub, & Bourdreau, 2000). The adequacy of the model was assessed by goodness-of-fit tests and the examination of the standardised path coefficients. A nested model comparison analysis was also performed using the chi-square difference ($\chi^2_{\text{difference}}$) test to explore the mediating role of affective reactions to crowding on the relationships among crowding, passenger density, and the two outcomes.

## 4. Results

### 4.1. Exploratory factor analyses

Using Pearson correlations, all items significantly correlated at least .28 with the passenger density item. All items are also significantly correlated at least .314 with at least one other item, suggesting reasonable factorability. Exploratory factor analyses were conducted using three separate principal axis factoring (PAF) extraction methods with varimax rotation to assess the dimensionality of these subscales. Separate analyses were employed because, as previously stated in Sections 1 and 2, the literature suggests that the experience of crowding is made up of three distinctive theoretical components. A summary of the PAF results is presented in Table 1.

The results show that all items correlate at least .50 with at least one other item in their hypothesised latent factors. Each PAF suggests that one factor that could be extracted is related to its hypothesised component. Overall, the loadings for each subscale are consistent and none of the items are loaded highly on more than one factor. The Kaiser–Meyer–Olkin index of sampling adequacy is .932, .961, and .819, respectively, while Bartlett's measures are all significant, both supporting the factorability of the correlation matrices. Altogether, 73.73% of total variance was accounted for by the evaluation of the psychosocial aspects subscale; 78.92% of total variance was accounted for by the affective reactions to the crowded situation subscale; and 71.93% of total variance was accounted for by the evaluation of the ambient environment subscale.

### 4.2. Internal reliability

The Cronbach's alpha value for each subscale exceeds the .70 level recommended by Nunnally and Bernstein (1994): respondents' evaluation of the psychosocial aspects of the crowded situation ($\alpha = .94$) and of its ambient environment ($\alpha = .87$), and their affective reactions to the crowded situation ($\alpha = .97$); indicating fairly strong scale reliability and internal consistency. No substantial increases in alpha for any of the subscales could have been achieved by eliminating more items.

### 4.3. Confirmatory factor analysis

#### 4.3.1. Validity and reliability of the measurement model

First, all factor loadings were significant ($p < .001$), demonstrating that the items for each latent variable reflect a single underlying construct. Second, an analysis of indicator reliability was performed on all items. The reliability of an indicator or observed variable is defined as the square of the correlation between a latent factor and that indicator. This value represents the percentage of variation in the indicator that is explained by the factor or $R^2$ (Hatcher, 1994). According to

**Table 1**
Factor loadings based on three PAF analyses ($n = 525$).

| Measure | Items | Factor loading[a] | Factor loading[a] | Factor loading[a] | Cronbach's alpha | Total variance explained (%) |
|---|---|---|---|---|---|---|
| *Subscale 1* | Unpleasant | .878 | | | .940 | 73.73 |
| Evaluation of the psychosocial aspects of the crowded situation | Disturbing | .877 | | | | |
| | Cluttered | .852 | | | | |
| | Chaotic | .834 | | | | |
| | Dense | .823 | | | | |
| | Disorderly | .817 | | | | |
| | Confining | .746 | | | | |
| *Subscale 2* | Irritable | | .903 | | .966 | 78.92 |
| Affective reactions to the crowded situation | Frustrated | | .902 | | | |
| | Tensed | | .895 | | | |
| | Distracted | | .884 | | | |
| | Stressful | | .874 | | | |
| | Hindered | | .864 | | | |
| | Restricted | | .860 | | | |
| | Uncomfortable | | .855 | | | |
| | Squashed | | .822 | | | |
| *Subscale 3* | Hot | | | .828 | .870 | 71.93 |
| Evaluation of the ambient environment | Smelly | | | .816 | | |
| | Stuffy | | | .794 | | |
| | Noisy | | | .726 | | |

Factor loadings <.40 are suppressed.
[a] Factor loadings presented are the unrotated factor matrix because only one factor was extracted for each subscale and the solutions cannot be rotated.

Fornell and Larcker (1981), indicator reliability should capture 50% of the variation in the indicator (i.e. should be greater than .50).

A summary of these results is depicted in Table 2 below. From Table 2, we can see that the indicator reliabilities for affective reactions to the crowded situation items were all above .686, while the indicator reliabilities for the evaluation of the psychosocial aspects subscale were all above the recommended range of .50. Regarding the range of indicator reliabilities for the evaluation of the ambient environment subscale, three items have relatively high reliabilities (.60 and above); however, one item (*Noisy*) has relatively low reliability (.469). Although the latter's result suggests that the validity of the item may be questionable, it was decided to retain this item in the model because it represented a key descriptor in this subscale. Furthermore, removing this item from the scale decreased the Cronbach's alpha for the scale to .854 from .870.

Third, composite reliability index was calculated to assess construct validity using the following equation (Hatcher, 1994):

$$\text{Composite reliability} = \sum [L_i]^2 \Big/ \left[\sum L_i\right]^2 + \sum \text{Var }(E_i)$$

where $L_i$ = the standardised factor loadings for that factor and Var ($E_i$) = the error variance associated with the individual indicator variables.

In general, composite reliability should be above .70, with .60 considered minimally acceptable (Hatcher, 1994; Bagozzi & Yi, 1988). As shown in Table 2, all subscales had composite reliability coefficients exceeding .83, indicating that the reliabilities of all subscales were acceptable.

Fourth, variance extracted estimates were computed using the following equation (Hatcher, 1994):

$$\text{Variance extracted} = \sum L_i^2 \Big/ \sum L_i^2 + \sum \text{Var }(E_i)$$

where $L_i$ = the standardised factor loadings for that factor and Var ($E_i$) = the error variance associated with the individual indicator variables.

These estimates assess the amount of variance that is explained by an underlying factor in relation to the amount of variance due to measurement error (Hatcher, 1994). Values of .50 or larger are considered desirable because estimates lower than .50 indicate that variance due to measurement error is greater than the variance captured by the latent factor (Hatcher, 1994; Fornell & Larcker, 1981). As seen in Table 2, the variance extracted estimates all exceed the .50 criteria, further suggesting the reliability and validity of all latent factors.

Fifth, convergent validity was determined by examining the statistical significance of the factor loadings of the indicators of each latent factor. Convergent validity is present when different instruments are used to measure the same construct, and scores from these different instruments are strongly correlated (Hatcher, 1994). For a probability of .05, the critical ratio should be higher than 1.96 and the standardised factor loadings of each indicator in the latent variable on which it saturates should be higher than .40 (Hair, Black, Babin, Anderson, & Tatham, 2006). Furthermore, Anderson and Gerbing (1988) suggest

**Table 2**
Psychometric properties of the measurement model.

| Constructs and indicators | Std. factor loading | Critical ratio and P: CR | P | Error variance[b] | Reliability: Indicator reliability[c] | Composite reliability | Cronbach's alpha α | Average extracted variance |
|---|---|---|---|---|---|---|---|---|
| *Evaluation of psychosocial aspects of the crowded situation* | | | | | | .940 | .940 | .691 |
| Unpleasant | .872 | | | .240 | .760 | | | |
| Cluttered | .856 | 26.879 | *** | .267 | .733 | | | |
| Disturbing | .899 | 29.753 | *** | .192 | .808 | | | |
| Chaotic | .848 | 26.425 | *** | .281 | .719 | | | |
| Confining | .726 | 20.241 | *** | .473 | .527 | | | |
| Disorderly | .802 | 23.811 | *** | .357 | .643 | | | |
| Dense | .805 | 24.001 | *** | .352 | .648 | | | |
| *Affective reactions to the crowded situation* | | | | | | .967 | .966 | .764 |
| Irritable | .903 | | | .185 | .815 | | | |
| Stressful | .876 | 30.851 | *** | .233 | .767 | | | |
| Hindered | .868 | 30.170 | *** | .247 | .753 | | | |
| Restricted | .857 | 29.307 | *** | .266 | .734 | | | |
| Frustrated | .899 | 33.008 | *** | .192 | .808 | | | |
| Distracted | .881 | 31.343 | *** | .224 | .776 | | | |
| Uncomfortable | .857 | 29.286 | *** | .266 | .734 | | | |
| Tensed | .894 | 32.453 | *** | .201 | .799 | | | |
| Squashed | .828 | 27.192 | *** | .314 | .686 | | | |
| *Evaluation of the ambient environment* | | | | | | .866 | .870 | .619 |
| Hot | .820 | | | .328 | .672 | | | |
| Noisy | .685 | 16.888 | *** | .531 | .469 | | | |
| Smelly | .775 | 19.887 | *** | .399 | .601 | | | |
| Stuffy | .857 | 22.820 | *** | .266 | .735 | | | |

[a] The first item for each scale was set to 1.00. All factor loadings were significant at $p < .001$.
[b] Error variance was calculated as 1 minus the indicator reliability.
[c] Indicator reliability scores were calculated as the square of the standardised factor loading.

that the parameter estimates should be high in value and *t*-values should be statistically significant, indicating that all indicators are effectively measuring the same construct. As shown in Table 2, all standardised loadings and critical ratios exceed the recommended minimum levels of .40 and 1.96, respectively. The measures in the resulting model indicate that convergent validity was obtained, with each measure being significantly related to its underlying factor.

Sixth, the discriminant validity (the degree to which two conceptually similar constructs are distinct) of the crowding subscales was determined in two ways. First, following suggestions by Kline (2011), the correlations between the constructs were examined. If the correlations are below .85, thus not highly correlated, then it can be assumed that they represent different constructs (Kline, 2011). As shown in Table 3, all correlations between the latent factors are below .85, providing evidence of discriminant validity.

Another way to measure discriminant validity is by comparing the variance extracted estimates for the factors with the squared correlations between the factors (Hatcher, 1994). Discriminant validity is demonstrated if all variance extracted estimates are greater than the squared correlation (Hatcher, 1994). As can be seen in Table 3, the variance extracted estimate was .691 for the evaluation of psychosocial aspects subscale and .619 for the evaluation of the ambient environment subscale. As the variance extracted estimates are greater than the square of the factor correlation (.493), the test supports the discriminant validity of these two factors. Examinations of other variance extracted estimates and squared correlation coefficients also generally support the discriminant validity of the components within the model.

In view of these data, findings from the exploratory and confirmatory factor analyses supported the hypothesised three-factor structure of the hypothesised model. All subscales demonstrated excellent internal consistency and construct validity as well as good convergent and discriminant validity of both the indicator items and their corresponding latent factors.

*4.3.2. Model estimation: goodness-of-fit indices*

Table 4 presents the goodness-of-fit results for the hypothesised three-factor measurement model. Although the model yielded a significant chi-square, $\chi^2$ ($n = 525$, df = 167) = 651.051, $p = .001$, a result that is not uncommon with large sample sizes (Byrne, 2009; Hooper et al., 2008), the other indices meet the criteria for adequacy of fit (SRMR = .034; CFI = .955; PNFI = .826; and RMSEA = .074; AIC = 737.51). These analyses thus further support and strengthen the research's hypotheses on the adequacy of the crowding measures developed in the present study.

**Table 3**
Discriminant validity among the latent factors/components.

| Latent factors/components | | 1 | 2 | 3 |
|---|---|---|---|---|
| 1 | Evaluation of the psychosocial aspects of the crowded situation | **.691** | *.714* | *.493* |
| 2 | Affective reactions to the crowded situation | .845 | **.764** | *.604* |
| 3 | Evaluation of the ambient environment | .702 | .777 | **.619** |

The diagonal entries (**in bold**) represent the average amount of extracted variance for each construct.
The Pearson correlations between constructs are shown in the lower triangle.
The variances shared between constructs are shown (*in italics*) in the upper triangle (calculated as the squares of correlations between constructs).
All correlations are significant at the $p < .001$ level.

**Table 4**
Fit indices for the hypothesised three-factor model.

| Fit index | Recommended criteria | Value |
|---|---|---|
| $\chi^2$ | • Chi-square value should not be significant if there is a good model fit (Byrne, 2009) | 651.051 |
| df ($p$) | | 167 (.001) |
| SRMR | • Good model fit: SRMR less than .05 (Byrne, 2009) | .034 |
| | • Adequate model fit: SRMR less than or equal to .08 (Hu & Bentler, 1999) | |
| CFI | • Good model fit: CFI equal or above .95 (Byrne, 2009; Hu & Bentler, 1999) | .955 |
| | • Adequate model fit: CFI less than or equal to .90 (Bentler & Bonett, 1980; Bollen, 1989) | |
| PNFI | • Recommended threshold of .60 (Marcoulides & Hershberger, 1997). | .826 |
| RMSEA | • Good model fit: RMSEA less than or equal to .05 (Byrne, 2009). | .074 |
| | • Adequate model fit: RMSEA less than or equal to .08 (Schumacker & Lomax, 2004; Byrne, 2009; Browne & Cudeck, 1992). | |
| AIC | • Good model fit: lower AIC value (Byrne, 2009). | 737.051 |

SRMR = Standardised Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; PNFI = Parsimonious Normed-Fit Index; AIC = Akaike's Information Criterion.

### 4.4. Comparisons of models

To further test the adequacy and validity of the hypothesised model, the three-factor model was tested against an independence model (no underlying relationships amongst factors), a one-factor model (where all items were collapsed into one global factor), and a series of two-factor models in which all possible combinations of the latent factors were considered. The three possible two-factor model combinations reflect the collapsing of (1) the evaluation of psychosocial aspects of the crowded situation with evaluation of the ambient environment; (2) the evaluation of psychosocial aspects of the crowded situation with affective reactions; and (3) the evaluation of the ambient environment with affective reactions. However, only the affective reactions versus the combined evaluation of psychosocial aspects of the crowded situation and evaluation of the ambient environment model yielded meaningful solutions. Therefore, the other two possible two-factor models were rejected on the basis of all fit indices criteria.

Fit indices for all models tested are presented in Table 5. Almost all fit indices for testing the independence model did not meet the recommended criteria, thus the model is easily rejected. The one-factor model provided a worse fit for the data than the two-factor model. On the other hand, the CFI value, which exceeds .95, and the SRMR value for the three-factor model, which is less than .05, were sufficient to allow it to be accepted. Furthermore, the RMSEA of .074 for the three-factor model indicates that the model is an adequate fit for the data (Byrne, 2009). These findings therefore suggest that the three-factor model provides the best fit to the data in comparison to other models. This conclusion is further supported by the best parsimonious fit (PNFI = .826) and the lowest Akaike's information criterion value (AIC = 737.051), which indicate that the three-factor model is better-fitting and more parsimonious than the other models tested.

### 4.5. Structural model analysis

#### 4.5.1. Model estimation: goodness-of-fit indices

Good support was found for the structural model: (SRMR = .029; CFI = .993; PNFI = .462; RMSEA = .058; and AIC = 47.306); although a significant chi-square was observed: [$\chi^2$ ($n$ = 525, df = 7) = 19.306, $p$ = .007]. Fig. 3 presents these results.

**Table 5**
Fit indices for each of the models tested.

| Model | $\chi^2$ (df) | SRMR (<.05) | CFI (⩾.95) | PNFI (⩾.60) | RMSEA (<.08) | AIC (lowest) |
|---|---|---|---|---|---|---|
| Independence model | 1857.379* (170) | .470 | .842 | .742 | .138 | 1937.379 |
| One-factor model | 1361.690* (170) | .052 | .888 | .783 | .116 | 1441.690 |
| Two-factor model | 1009.623* (169) | .049 | .921 | .807 | .097 | 1091.623 |
| Three-factor model | 651.051* (167) | .034 | .955 | .826 | .074 | 737.051 |

* Significant at $p < .001$ level.

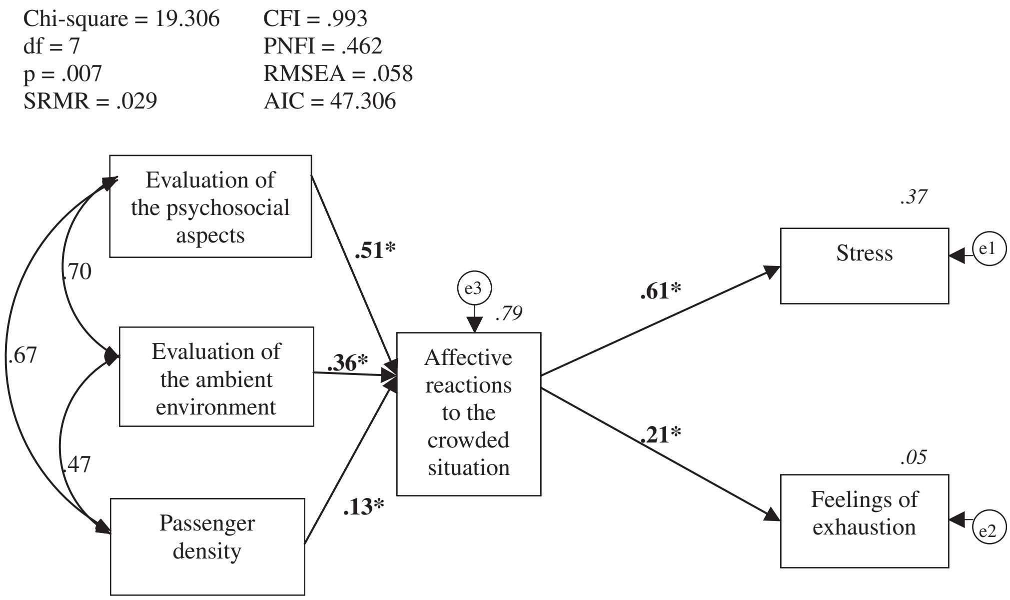


**Fig. 3.** Results for the hypothesised structural model.

In order to test H1 to H5, the statistical significance of all causal paths in the model was examined. The results showed that affective reactions to the crowded situation are significantly predicted by commuters' evaluation of the psychosocial aspects of the crowded situation ($\beta = .51$, $p < .001$), by their evaluation of the ambient environment ($\beta = .36$, $p < .001$), and by passenger density ($\beta = .13$, $p < .001$). 79% of the variance in the affective reactions to the crowded situation is accounted for by the combination of these three predictors. The affective reactions to crowding, in turn, significantly predict experience of stress ($\beta = .61$, $p < .001$) and feelings of exhaustion ($\beta = .21$, $p < .001$). This suggests that the more unpleasantly crowded the commuters feel, the more stress and feelings of exhaustion they experience. Altogether, 37% of the variance in the experience of stress and 5% of the variance in feelings of exhaustion was explained by affective reaction to crowding.

*4.5.2. Nested model comparison*

To more closely examine the affective reactions to the crowded situation's role within the model, a nested model comparison analysis with a baseline model was conducted (see Fig. 4 below).

The posited model contains three nested models. The first is the unconstrained model where all paths among the variables were free to vary (Model 1). Second, a direct path model (Model 2) was assessed. This model incorporates direct paths from (1) the evaluation of the psychosocial aspects of the crowded situation to the outcomes (stress and feelings of exhaustion), (2) the evaluation of the ambient environment to the outcomes, and (3) rated passenger density to the outcomes, while simultaneously constraining all paths to/from affective reactions to crowding to 0. In essence, this direct path model tests for the direct effects of passenger density as well as the evaluation of the psychosocial aspects of the crowded situation and of its ambient environment on the outcomes without the effect of the affective reactions to the crowded situations component. Third, the hypothesised model, wherein all direct paths from the three predictors to the outcomes were constrained to 0, was estimated (Model 3). The chi-square difference ($\chi^2_{\text{difference}}$) test was used to compare these models as they are hierarchical models based on the same dataset (Kline, 2011). The presence of a significant $\chi^2_{\text{difference}}$ value indicates that the model with more paths explains the data better (Kline, 2011).

The comparison of Models 1 and 2 indicates that the unconstrained model provides a better fit than the direct path model ($\chi^2_{\text{difference}} = 855.649$, df = 5, $p < .001$). However, the chi-square difference for Models 1 and 3 was not statistically significant ($\chi^2_{\text{difference}} = 8.191$, df = 6, $p = .224$), indicating that Model 3 explains the data equally well compared to the unconstrained Model 1. This result suggests that Model 3 is the more parsimonious model than the unconstrained model, and is therefore preferred. In addition, all fit indices for the hypothesised model (Model 3) improved and the improvement was significant against the unconstrained model, with SRMR, CFI, and PNFI (.029, .993, and .462, respectively), and RMSEA (.058) all at acceptable levels. The hypothesised model also obtains the lowest AIC value (47.306) relative to other models, further confirming its model parsimony. Table 6 provides the comparison of fit indices of these models.

An examination of path coefficients among variables in the unconstrained model (Model 1) shows that the evaluation of the psychosocial aspects of the crowded situation and of its ambient environment and passenger density have significant effects on affective reactions to the crowded situation as predicted. However, the direct paths linking these predictors to stress and feelings of exhaustion were all statistically non-significant (see Table 7). These results demonstrate that neither

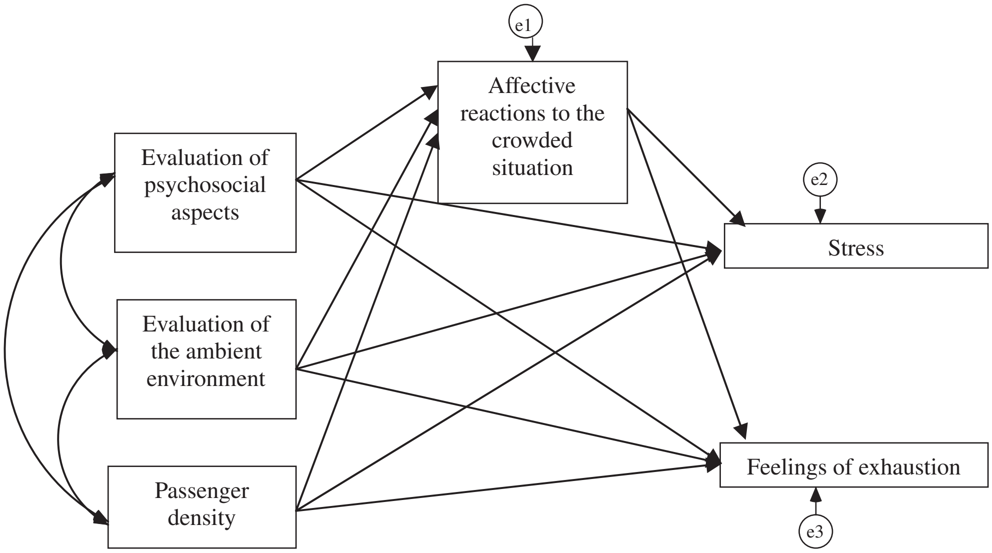


**Fig. 4.** Path model for nested model comparison.

**Table 6**
Fit indices for the nested model comparison.

| Model | $\chi^2$ (df) | SRMR (<.05) | CFI (⩾.95) | PNFI (⩾.60) | RMSEA (<.08) | AIC (lowest) |
|---|---|---|---|---|---|---|
| Model 1 (unconstrained) | 11.115** (1) | .024 | .994 | .066 | .139 | 51.115 |
| Model 2 (direct path) | 866.764** (6) | .320 | .507 | .203 | .523 | 896.764 |
| Model 3 (hypothesised) | 19.306** (7) | .029 | .993 | .462 | .058 | 47.306 |

** Significant at $p < .001$ level.

**Table 7**
Regression weights[a] of the baseline/unconstrained model.

| Parameter | | | Unstandardised solutions | | Standardised solutions | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $\beta$ | SE | $\beta$ | SE | Bias-corrected (90% C.I.) | $p$ value |
| Affective reactions | ← | Evaluation (psychosocial) | .71 | .06 | .51 | .04 | (.44, .57) | .002** |
| Affective reactions | ← | Ambient environment | .89 | .08 | .36 | .03 | (.30, .41) | .003** |
| Affective reactions | ← | Rated passenger density | 1.31 | .29 | .13 | .03 | (.08, .18) | .001** |
| Stress | ← | Evaluation (psychosocial) | .02 | .05 | .03 | .07 | (−.08, .16) | .643 |
| Feelings of exhaustion | ← | Evaluation (psychosocial) | .05 | .08 | .05 | .08 | (−.07, .18) | .528 |
| Stress | ← | Ambient environment | .09 | .06 | .09 | .06 | (.00, .19) | .100 |
| Feelings of exhaustion | ← | Ambient environment | .08 | .12 | .05 | .07 | (−.06, .17) | .434 |
| Stress | ← | Rated passenger density | .32 | .22 | .07 | .05 | (−.01, .15) | .146 |
| Feelings of exhaustion | ← | Rated passenger density | −.58 | .432 | −.08 | .06 | (−.17, .03) | .234 |
| Stress | ← | Affective reactions | .21 | .03 | .47 | .07 | (.34, .58) | .002** |
| Feelings of exhaustion | ← | Affective reactions | .13 | .07 | .18 | .09 | (.02, .33) | .072 |

Squared multiple correlation coefficients ($R^2$) for: affective reactions to the crowded situation = .79; for stress = .38; and for feelings of exhaustion = .05.
[a] Results based on bootstrap ML analysis procedures.
** Significant at $p < .001$ level.

evaluation of the psychosocial aspects of the crowded situation, evaluation of the ambient environment, nor rated passenger density, were directly related to the experience of stress and feelings of exhaustion.

Following these results, it can be further suggested that there is an interplay between the different components of crowding experience and passenger density, such that commuters' evaluations of the psychosocial aspects of the crowded situation and of its ambient environment and their ratings of passenger density are related to the experience of stress and feelings of exhaustion, being mediated by their affective feelings of crowdedness.

## 5. Discussion

This study aimed at addressing two issues: (1) the development of a measure that captures the dimensionality and measurement of rail passenger crowding, and (2) its relationship to the experience of stress and feelings of exhaustion. The adequacy and validity of the proposed instrument was tested on 525 commuters using rigorous statistical analyses in order to ensure that it possesses desirable psychometric properties. An iterative refinement of the dataset by means of factor analyses, reliability and validity analyses as well as structural equation modelling resulted in the development of a three-factor instrument that was considered to reflect potentially important components of the crowding experience. Derived constructs, identified as the evaluation of the psychosocial aspects of the crowded situation, the evaluation of the ambient environment of the crowded situation, and affective reactions to the crowded situation, were found to demonstrate good model fit and possess sufficiently good construct, convergent, and discriminant validity values.

The instrument developed in this thesis is concordant with the work of Kalb and Keating (1981), which suggests a distinction between the crowding experience, the essentially psychosocial characteristics of the crowded situation, and those of its ambient environment. Consisting of 20 items, the present instrument is both multidimensional and sufficiently robust to represent the underlying theoretical constructs of the crowding experience. Further, the instrument's factorial structure confirms the value of considering descriptive-psychosocial, ambient, and affective components in assessing how individuals conceptualise their crowding experiences.

The descriptive-psychosocial component of crowding experience has already been established in detail through research of environmental stressors (Evans & Cohen, 2004; Bell et al., 2001). This is reflected in the first subscale of the present instrument which characterises negative evaluations of the density level experienced by individuals such as dense, disorderly, confining, chaotic, cluttered, disturbing, and unpleasant. The ambient component associated with the evaluation of the physical environment of the crowded setting was reflected in four items: hot, noisy, stuffy, and smelly. Research has established that heat, odour, and discomfort are associated with the feeling of crowdedness (Braun & Parsons, 1991) and may contribute to aggressive behaviour and negative interpersonal responses (Griffit & Veitch, 1971). Meanwhile, the affective component of the instrument focuses on the individual's experiential feeling state in crowded situations such as behavioural constraint, interference, and stress. Researchers such as Stokols (1972), Kalb and Keating (1981), and Cox et al. (2006), among others, have reported that the individual's perception of the aversive experiential state or feeling of crowdedness is linked to constraint, distraction, discomfort, and stress.

With regard to the experience of stress and feelings of exhaustion, the proposed rail passenger crowding instrument produced expected correlations with the outcome measures and a good model fit of the data with the overall theoretical structure. The nested model comparison results suggest that the links between rail passenger crowding and the outcomes do not appear as a simple, direct relationship, but are mediated through affective reactions to the crowded situation. These findings, in particular, challenge the popularly assumed link between physical density and negative outcomes and suggest that passengers could still experience stress and exhaustion even under what appears to be low-density conditions. Indeed, this evidence supports the proposition that density per se does not necessarily lead to the experience of crowding or to the negative effects associated with crowding (Stokols, 1972).

One important implication of the present study is that the findings offer an opportunity to challenge the prevailing passenger density and loading measures used in the rail industry. Up until recently, the literature dealing with crowding in this setting predominantly focuses on passenger density, and the pathways that may link the subjective aspect of crowding with outcomes have not been fully explored. The findings of the present work to some extent challenge this norm and make the case for the role and significance of the psychological components in assessing crowding experiences among passengers. In particular, the present work argues that individuals' crowding experience is an essentially subjective, psychological phenomenon and is better represented as a multidimensional construct rather than as a single, unidimensional experience. However, being multidimensional about the experience of crowding is not about adding up the different components and obtaining a grand total score of the magnitude of the crowding experience. Rather, it is about explicitly recognising that there are important components of crowding that cannot be adequately represented by a single global rating or score. For a more accurate measurement and conceptualisation of the crowding experience, it is therefore recommended that future studies should use this instrument (or something similar to it) in order to better understand this phenomenon.

Although the instrument advanced in this study contributes constructively to the literature, some possible enhancements that could improve its utility and potentiality in assessing the experience of crowding and its various outcomes are suggested. First, it is recommended to compare the proposed instrument with other crowding measures, as studies with different subjective crowding measures could provide better support for its criterion-related validity. Second, the study data were self-reported, which raises the possibility of participant bias, social desirability, demand characteristics, and response sets. Such possibilities may, in turn, affect the validity of the findings. Thus, future research may also incorporate other objective or independent measures to supplement the subjective evaluation of crowding and enhance the interpretation of the findings. Third, studies have established that perceived control and predictability are two potential psychological factors influencing the relationship between crowding and stress (Cohen & Sherrod, 1978; Valins & Baum, 1973; Baum & Koman, 1976). Prior research specifically in the rail literature also reveals that a lack of control over commuting conditions (e.g. Singer et al., 1978; Lundberg, 1976), a decreased sense of predictability of the commute (e.g. Evans, Wener, & Phillips, 2002), intrusion into personal space on public transport (e.g. Evans & Wener, 2007), and increased commuting duration (e.g. Evans & Wener, 2006) have all been found to play important moderating or mediating roles in examining the nature and effects of rail passenger crowding. However, the roles of these variables are not specifically addressed in the current study. Subsequent analysis is being conducted to identify their potential role on this issue. Finally, while the sample in this study was representative of those workers who frequently commute to work using either the light rail transit or commuter rail services, the same findings obtained may not hold true for moderate or occasional rail users, who have much less experience of stress associated with rail crowding. Therefore, one should exercise caution not to generalise the present results beyond the specific population used or the specific variables employed, at least until further research explores a similar phenomenon in other research contexts.

In conclusion, the present study has presented evidence that the proposed instrument has good psychometric properties and is a promising tool for the measurement of crowding experience among rail users. The conceptual and practical utility of focusing on subjective experience of crowding rather than an objective measure of density was demonstrated and provided adequate support for a mediational model, in which affective reactions to the crowded situation play a role in the experience of stress and feelings of exhaustion among commuters. In doing so, the results clarify that there is an interplay among the different dimensions of crowding experience, passenger density, and the negative outcomes.

## Acknowledgements

This study was completed as part of Ph.D. research by the first author under the financial sponsorships of the Ministry of Higher Education Malaysia and the International Islamic University Malaysia. It has been reviewed and approved by the Institute of Work, Health and Organisations' Ethics Committee (Ref. No.: November 7th, 2008) and by the Economic Planning Unit (EPU), Prime Minister's Department, Malaysia (Ref. No.: UPE: 40/200/19/2410). The authors gratefully acknowledge TRANSIT.org, Ministry of Transport Malaysia, Department of Railways Malaysia, Keretapi Tanah Melayu Bhd (KTMB) as well as Rangkaian Pengangkutan Deras Sdn Bhd (RapidKL) for providing support to this research. The authors are also especially grateful to two anonymous reviewers for providing useful comments and insightful suggestions. Finally, thanks are due to all enumerators and respondents for their assistance in this research.

## References


Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411–423.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*, 74–94.

Baum, A., & Koman, S. (1976). Differential response to anticipated crowding: Psychological effects of social and spatial density. *Journal of Personality and Social Psychology, 34*, 526–536.

Baum, A., & Paulus, P. B. (1987). Crowding. In D. Stokols & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 533–570). New York: Wiley.

Bell, P., Green, T., Fisher, J., & Baum, A. (2001). *Environmental psychology*. USA: Wadsworth Group/Thomson Learning.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Bollen, K. A. (1989). *Structural equations with latent variables. Wiley series in probability and mathematical statistics*. New York: Wiley.

Braun, T. L., & Parsons, K. C. (1991). Human thermal responses in crowds. In E. J. Lovesey (Ed.), *Contemporary ergonomics* (pp. 190–195). London: Taylor & Francis.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230–258.

Buckley, F., & O'Regan, B. (2004). *The psychological effects of commuting in Dublin*. The Learning, innovation and knowledge. Research centre working paper series, 07-04. <http://doras.dcu.ie/2407/1/wp0704.pdf>.

Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic concepts, application, and programming* (2nd ed.). New York: Routledge Academic.

Cassidy, T. (1992). Commuting-related stress: Consequences and implications. *Employee Counseling Today, 4*, 15–21.

Cohen, S., & Sherrod, D. R. (1978). When density matters: Environmental control as a determinant of crowding effects in laboratory and residential settings. *Journal of Population, 1*, 189–202.

Cox, T., & Gotts, G. (1987). *The General Well-Being Questionnaire (GWBQ) manual*. Nottingham: Department of Psychology, University of Nottingham – Stress Research Unit.

Cox, T., Houdmont, J., & Griffiths, A. (2006). Rail passenger crowding, stress, health and safety in Britain. *Transportation Research Part A – Policy and Practice, 40*, 244–258.

Dodgson, J., Kelso, E., Van der Veer, J. P., Skene, R., & Paredes, D. (2002). *The tube: Moving on final report*. London: National Economic Research Associates (NERA). <http://legacy.london.gov.uk/assembly/reports/transport/tube_priorities/NERA_final_report.pdf>.

Evans, G. W., & Cohen, S. (2004). Environmental stress. In C. D. Spielberger (Ed.). *Encyclopedia of applied psychology* (Vol. 3, pp. 815–824). New York: Elsevier. <http://www.psy.cmu.edu/~scohen/evansCohenChap04.pdf>.

Evans, G. W., Lepore, S. J., & Mata-Allen, K. (2000). Cross-cultural differences in tolerance for crowding: Fact or fiction? *Journal of Personality and Social Psychology, 79*, 204–210.

Evans, G. W., & Wener, R. E. (2006). Rail commuting duration and passenger stress. *Health Psychology, 25*, 408–412.

Evans, G. W., & Wener, R. E. (2007). Crowding and personal space invasion on the train: Please don't make me sit in the middle. *Journal of Environmental Psychology, 27*, 90–94.

Evans, G. W., Wener, R. E., & Phillips, D. (2002). The morning rush hour: Predictability and commuter stress. *Environment and Behaviour, 34*, 521–530.

Fischer, C. S., Baldassare, M., & Ofshe, R. J. (1975). Crowding studies and urban life: A critical review. *Journal of the American Planning Association, 41*, 406–418.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable and measurement error. *Journal of Marketing Research, 18*, 39–50.

Freedman, J. L. (1975). *Crowding and behaviour*. San Francisco: W.H. Freeman and Company.

Fuller, T. D., Edwards, J. N., Sermsri, S., & Vorakitphokatorn, S. (1993). Housing, stress, and physical well-being: Evidence from Thailand. *Social Science and Medicine, 36*, 1417–1428.

Gefen, D., Straub, D., & Bourdreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*, 1–78 (Article 7).

Gotts, G., & Cox, T. (1988). *Stress and arousal checklist: A manual for its administration, scoring and interpretation*. Melbourne: Swinburne University Press.

Griffit, W., & Veitch, R. (1971). Hot and crowded: Influences of population density and temperature on interpersonal affective behavior. *Journal of Personality and Social Psychology, 17*, 92–98.

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, New Jersey: Pearson-Prentice Hall.

Harrell, G. D., Hutt, M. D., & Anderson, J. C. (1980). Path analysis of buyer behavior under conditions of crowding. *Journal of Marketing Research, 17*, 45–51.

Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc..

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*, 53–60.

Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.

Jain, A. K., Cox, T., & Houdmont, J. (2006). Passengers experience of rail travel. *Proceedings of the British Psychological Society, 14*, 48.

Kalb, L. S., & Keating, J. P. (1981). The measurement of perceived crowding. *Personality and Social Psychology Bulletin, 7*, 650–654.

Kaya, N., & Weber, M. J. (2003). Cross-cultural differences in the perception of crowding and privacy regulation: American and Turkish students. *Journal of Environmental Psychology, 23*, 301–309.

Kline, R. B. (2011). *Principles and practice of structural equation modelling* (3rd ed.). New York: The Guilford Press.

Lam, W. H. K., Cheung, C. Y., & Lam, C. F. (1999). A study of crowding effects at the Hong Kong light rail transit stations. *Transportation Research Part A – Policy and Practice, 33*, 401–415.

Lam, W. H. K., Cheung, C. Y., & Poon, Y. F. (1999). A study of passenger discomfort measures at the Hong Kong mass transit railway system. *Journal of Advanced Transportation, 33*, 389–399.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563–575.

Lundberg, U. (1976). Urban commuting: Crowdedness and catecholamine excretion. *Journal of Human Stress, 2*, 26–32.

Machleit, K. A., Kellaris, J. J., & Eroglu, S. A. (1994). Human and spatial dimensions of crowding perceptions in retail environments: A note on their measurement and effect on shoppers' satisfaction. *Marketing Letters, 5*, 183–194.

Marcoulides, G. A., & Hershberger, S. L. (1997). *Multivariate statistical methods*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mohd Mahudin, N. D. (2003). *Development of a Crowd Stress Index (CSI)*. Unpublished master's thesis, Loughborough University, United Kingdom.

Mohd Mahudin, N. D., Cox, T., & Griffiths, A. (2010). *The effects of rail passenger crowding on health and stress: A systematic review*, submitted for publication.

Nunnally, C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ping, T. L. (2005). *Does the survey response scale format matter?* Paper presented at the 47th annual conference of International Military of Testing Association (IMTA) 2005, Singapore. <http://www.internationalmta.org/Documents/2005/2005133P.pdf>.

Saris, W., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with construct-specific response options. *Survey Research Methods, 4*, 61–79.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modelling* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Singer, J., Lundberg, U., & Frankenhauser, M. (1978). Stress on the train: A study of urban commuting. In A. Baum, J. Singer, & S. Valins (Eds.). *Advances in environmental psychology, the urban environment* (Vol. 1, pp. 41–56). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stokols, D. (1972). On the distinction between density and crowding: some implications for further research. *Psychological Review, 79*, 275–277.

Thomas, L. J., Rhind, D. J. A., & Robinson, K. J. (2006). Rail passenger perceptions of risk and safety and priorities for improvement. *Cognition, Technology and Work, 8*, 67–75.

Turner, S., Corbett, E., O'Hara, R., & White, J. (2004). *Health and safety effects of rail crowding – Hazard identification*. HSL report RAS/04/12. <http://www.rssb.co.uk/SiteCollectionDocuments/pdf/reports/Research/T307_rpt_final.pdf>.

UK Department for Transport (2009). *Transport trends 2009: Section 3 public transport*. <http://www2.dft.gov.uk/pgr/statistics/datatablespublications/trends/current/section3pubtran.pdf>.

UK Rail Safety and Standards Board: RSSB (2004). *Crowd management at stations: A good practice guide*. London: Rail Safety and Standards Board. <http://www.rssb.co.uk/SiteCollectionDocuments/pdf/CMaS%20-%20A%20good%20practice%20guide.pdf>.

Valins, S., & Baum, A. (1973). Residential group size, social interaction and crowding. *Environment and Behavior, 5*, 421–439.

Wener, R. E., & Evans, G. W. (2004). *The impact of mode and mode transfer on commuter stress: The Montclair Connection*. Final report. No. FHWA-NJ-2004-005. <http://www.utrc2.org/research/assets/74/commuterstress2-report1.pdf>.

Yusoff, M. H. (2008). Central Region Manager. *KTM Komuter survey 2005* (personal communication, 14th September 2008).