

## Research Article

# Analysis of Work-Zone Crashes Using the Ordered Probit Model with Factor Analysis in Egypt

Kairan Zhang,<sup>1</sup> Mohamed Hassan ,<sup>2</sup> Mahama Yahaya,<sup>2</sup> and Shupeng Yang<sup>2</sup>

<sup>1</sup>Professor, School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China

<sup>2</sup>School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China

Correspondence should be addressed to Mohamed Hassan; abo\_heseny@yahoo.com

Received 16 May 2018; Revised 2 September 2018; Accepted 17 September 2018; Published 1 October 2018

Academic Editor: Constantinos Antoniou

Copyright © 2018 Kairan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Work-zones, due to their nature, are predisposed to hazardous situations. This is a consequence of conducting construction work within the vicinity of, or near, vehicular traffic. The exposure to danger underscores the need for proper understanding of the occurrence of work-zone crashes, as well as the risk factors associated with them. This paper aims mainly to develop a hybrid approach that combines a factor analysis method and an ordered probit model to carry out a comprehensive analysis of work-zone crashes. The paper presents an analysis of work-zone data crashes from 2010 to 2015 that occurred in Egyptian long-term highway maintenance and rehabilitation projects. Factor analysis is used to identify the main and common factors that influence work-zone crashes and to calculate the weight of every factor. The ordered probit model is developed, based on the results of factor analysis scores, to examine the contribution of common factors in the severity of work-zones. The most influential factors that have contributed to work-zone crashes are weather condition, number of lane closures, type of surface construction, road character, day of week, and so forth. In addition, the results indicated that four common factors are significantly affecting the severity of work-zone crashes in Egypt.

## 1. Introduction

For the last few years, Egypt has recorded such a large number of traffic crashes that “death on the road” is frequently seen on signs as a way of describing highway safety in Egypt. The fatalities from road traffic crashes for 2013 were 10,466, which had a great impact on the emerging economy [1]. As highway infrastructure ages, there is a growing need for regular maintenance and rehabilitation. This leads to an increase in the number of work-zones on highways, with the attendant disruption and interruption of traffic flow, resulting in traffic safety problems. Closing off highways to traffic, while maintenance and rehabilitation works are ongoing, is usually unrealistic. This leads to only a partial closure, with redirection of traffic into alternative lanes and routes. Scenarios like this are typical, creating traffic work-zones, resulting in constricted space for traffic. As a result of all this, the safety of work-zones occupies a place of extreme importance in Egypt and the Middle East in general. Therefore, the pressure on government and road

safety managers to institute measures that can reduce the frequency of road crashes and the severity of injuries is on the increase. However, no previous research about work-zone crashes in Egypt has been published. Thus, research work on crash severity may lead to the discovery of factors that influence changes in severity and that may be potentially beneficial in developing traffic controls, to minimize the incidence of severe highway crashes.

This paper mainly aims to develop a hybrid approach, which combines the factor analysis and ordered probit model to carry out a comprehensive analysis of work-zone crashes. The first stage, the factor analysis method, is used based on work-zone crash data elements to determine the weight of each factor to classify the main factors and identify the common factors. In the second stage, the severity model is developed based on common factor scores to determine the relative effects of each common factor. An ordered probit model is an appropriate model to identify the characteristics that contribute to crash severity, because crash severity, as a dependent variable, is ordered in nature [2–4].

## 2. Literature Review

Around the world, significant research effort had already been conducted in order to identify the factors and causes that have an influence on work-zone crashes with regard to its severity and frequency. Essential portion of the research revealed that part of an effective and efficient work-zone safety is utilizing statistical modeling techniques to determine the contributing factors associated with work-zone crashes.

Dissanayake and Lu [5] developed a set of sequential binary logistic regression models to predict the factors causing crashes. However, the methodology does not account for multivehicle collisions instigated by young drivers as well as the condition of the damaged vehicles. Abdel-Aty et al. [6] used the log-linear technique to examine the connection between driver age and crash characteristics. They concluded that injury severity was positively associated with the age of the driver. Furthermore, [7] conducted a study using a negative binomial regression technique to analyze cross-median crashes statistically. Two predictive models were developed, by considering three factors: median width, Annual Daily Traffic (ADT), and segment length. Donnell and Mason [8] developed logistic regression models, by using both the ordinal and nominal responses as a way of demonstrating crash severity. From the study, the ordinal response model worked better for cross-median crashes, whereas the nominal response gave a better outcome for median-barrier crashes. Osman et al. [9] conducted a study, which intended to assess the factors to investigate the contribution of several factors to severity of collisions in work-zones. The model showed that roadways class, crashes during weekends and evening times, and road alignment were significantly affecting the severity of work-zone crashes, whereas, in New Zealand, [10] concluded that the time period, heavy vehicle, and vulnerability of drivers were the factors which determined the severity of collisions in work-zones. Li and Bai [11] examined work-zone crash severity by using crash severity index models; they recommended that during work-zone safety inspection or work-zone planning crash severity index models should be used. Theofilatos et al. [12] applied meta-analyzed technique to examine the connection between crash frequency and work-zone length and duration. Yang et al. [13, 14] devised a crash frequency model and further evaluated the measurement errors in work-zone length, by using full Bayesian estimations rather than the NB model, to identify risk factors in work-zone safety. A study conducted by [15] used the negative binomial model to compute monthly crashes in the work-zone. From the study, they concluded that working at night resulted in minimal crashes, compared to daytime. However, they could not identify an acceptable relationship between injury frequency and presence. Chang and Mannering [16] studied the occupancy-crash-injury severity relationship and deduced that truck-involved crashes, rear-end crashes, and high-speed limits increased crash injury severity; whereas the logit model was used by [17] to analyze driver injury severity involved in a single-vehicle crash, the researchers concluded that several major factors increase driver injury severity. These included daylight, speed limit, and dry pavement as opposed to a slippery pavement, among others.

Khattak and Targa [18] and Khattak et al. [19, 20] applied the ordered probit models in predicting injury level for crashes at construction zones, involving trucks, to identify characteristics that contribute to the severity of accidents involving older drivers, while [2] applied this technique to show the similarities and differences in factors contributing to crash injury severity on roadway sections, signalized intersections, and toll plazas; he concluded that results produced by the model were the best, besides it is worth due to its simplicity. Thus, he recommended using the order probit approach to model driver injury severity. Additionally, the ordered probit model was used for crash severity level, and as a result, the study encouraged brief reporting of occurrences, as this made their retention and storage in crash databases easier [21]. On the same note, [4] used the same model in investigating the risk of different injury levels sustained in two-vehicle crashes and single-vehicle crashes, with the conclusion that pickups and sports vehicles were less safe than passenger cars. Also, the study concluded that males and younger drivers, in newer vehicles at lower speeds, suffered less severe injuries. The same technique was applied by [22] to check factors influencing the severity of injury faced by motor-vehicle occupants. According to the authors, rural areas recorded more serious crashes than in urban areas, and women were more likely to be involved in serious or fatal injuries crashes than men. Qi et al. [23] conducted a study on rear-end crashes in the work-zone, through the stepwise elimination selection method, for calibrating the ordered probit model. A limitation of the methodology is that it lacks vital information, such as various vehicle characteristics, external factors, such as light and weather conditions, and the drivers' age and sex, in their severity model. According to [24], surface conditions contributed to the severity of work-zone crashes. For instance, dry conditions were more dangerous and severe than wet ones. Furthermore, middle-aged drivers were more susceptible to work-zones crashes, according to [25].

Xu [26] applied decision tree analysis, correlation analysis, and cluster analysis as data mining methods to analyze traffic crash data. Sun et al. [27] developed the Analytic Hierarchy Process (AHP) considering causes and environment, road, and the influence of people on traffic crashes. Cheng [28] studied the correlation between traffic crashes and different factors were analyzed. The study concluded that traffic crashes had a significant correlation with drivers, number of vehicles, and population. Chen et al. [29] conducted a study using factor analysis to study the major factors contributing to road traffic crashes. The researchers concluded that the major factors in crashes depend on their weights, faulty behavior, driving experience, the condition of the vehicle, the purpose of the vehicle, and so on. On the same note, Yuan et al. [30] used the same technique to determine the weight of main factors through the factor score matrix.

This study is an attempt to bridge the gap in knowledge of the work-zone crashes in Egypt. To our knowledge, no previous research about work-zone crashes in Egypt has been published before. This study is among the first to investigate the factors affecting the severity of Egyptian highways work-zone crashes. Second, the results and conclusions of this

paper can provide insightful information for stakeholders in the planning and management of work-zones.

### 3. Data Collection

The focus of this paper rested on work-zone crashes recorded during highway maintenance and rehabilitation projects, on both urban and rural highway facilities, with long-term projects (duration greater than one year) with both day and night work. In Egypt, the Ministry of the Interior has a traffic department, with a research unit responsible for managing the database of national road crashes. A subsidiary of the Ministry of Transport, the General Authority for Roads, Bridges and Land Transport (GARBLT), also regularly collects crash data, especially those that occur on federal highways. Crash data of about 345 cases, spanning 2010 through 2015, was used in this paper. Crash variables extracted from the database for each incident included information on the driver, vehicle information, time, characteristics of the road, work-zone information, and environmental conditions. Six tables were used in presenting the data that was obtained. These included fault drivers table, environmental table, work-zone information table, road character table, vehicle and crash characteristics table, and crash severity table. Table 1 illustrates the frequency distribution and descriptive statistics of the influential factors.

### 4. Methodology

**4.1. Factor Analysis.** Factor analysis [31] is a method of statistics which aims at finding the most important factors that can cause an event, for a number of plausible reasons. It is a collection of various methods, used to study the reason behind underlying constructs that may, or can, influence the responses to the values of observed data. Its main aim is to determine the most important and common factors that are influencing a set of observed measures. In this study, we apply the principal component method to extract maximum variance from the set of work-zone crash variables and put them into common factors as index for the analysis. The general model of factor analysis, according to [32], is given by

$$X = AF + B \quad (1)$$

$$x_1 = a_{11}F_1 + a_{12}F_2 + a_{1m}F_m + \varepsilon_1$$

$$x_1 = a_{21}F_1 + a_{22}F_2 + a_{2m}F_m + \varepsilon_2 \quad (2)$$

$$x_1 = a_{p1}F_1 + a_{p2}F_2 + a_{pm}F_m + \varepsilon_p$$

Here  $X$  is the observable random vector (i.e., original observed variables);  $F$  is the common factor of  $X$ ,  $A$  is the coefficient of  $F$  (factor loading matrix),  $a_{ij}$  is the correlation coefficient,  $\varepsilon_i$  is the error factor, and  $B$  is the special factor of  $X$  (usually ignored when analyzing).

$a_{ij}$  determines the degree or strength of a linear relationship between the variables that are present in the rows

and columns of the matrix. The higher the correlation is, the stronger that relationship will be.

$$x_i = a_{i1}F_1 + a_{i2}F_2 + a_{im}F_m + \varepsilon_i \quad (3)$$

The process of factor analysis involves normalizing the  $X$  matrix so that its mean value is 0 and its variance is 1, under the assumption. Then we need to assume that  $F$  and  $\varepsilon_i$  are independent of each other. Here,  $X$  has  $m$  common factors and is known as the factor model. After that, the correlation coefficient matrix, i.e.,  $R = (rij)p * p$ , and its latent root, i.e.,  $\lambda$ , are obtained. The last step is to determine the number of factors that are common and relevant and add up the common variance, i.e.,  $h^2_i$ , and then eventually the common factors are obtained by rotating the loading matrix.

**4.2. Ordered Probit Model.** Ordered probit model is the other statistical modeling methodology that has been used in this paper. This model is usually built on the idea of a latent factor underlying the injury risk propensity from road crashes. These road crashes were studied to determine the observed ordinal fatal injury crash reasons [33]. The general specification of the model is given by [34]

$$z_i = \beta X_i + \varepsilon_i \quad (4)$$

Here  $z_i$  is the latent risk propensity for crash victim  $i$ ,  $\beta$  is the estimated vector of parameters, and  $X_i$  is the vector of observed nonrandom explanatory variables. It measures the various attributes of crash victims  $i$ , and  $\varepsilon_i$  is the random error term that follows a standard normal distribution.

Hence, its mean is 0, and the variance is 1 for  $\varepsilon_i$ .

The standard ordinary least square regression technique cannot be applied here, as the dependent variable of this model,  $z_i$ , is unobserved. It is also reasonable to assume that a high risk of severity is related to observed fatalities, denoted by  $y_i$ . This relationship has been described in the following equations [35, 36]:

$$\begin{aligned} y_i &= 1, & \text{if } z_i \leq \mu_1 \\ y_i &= k, & \text{if } \mu_{k-1} < z_i \leq \mu_k \\ y_i &= K & \text{if } z_i > \mu_{k-1} \end{aligned} \quad (5)$$

Here  $\mu = \{\mu_1 \dots \mu_k \dots \mu_{k-1}\}$  represent the threshold values for crashes that define  $y_i$ , corresponding to integer ordering, and  $K$  is the highest-ordered severity level. Now, it may be noted that the probability of a crash victim  $i$  facing a level  $k$  severity is always equal to the probability that  $z_i$  (i.e., the latent risk propensity) assumes a value which is between two fixed thresholds. It can also be explained as follows: if we are given the value of  $X_i$ , then the probability that the severity faced by crash victim  $i$  belongs to each severity level is given by

$$\begin{aligned} P(y = 1) &= \phi(-\beta X_i) \\ P(y = k) &= \phi(\mu_{k-1} - \beta X_i) - \phi(\mu_{k-2} - \beta X_i) \\ P(y = k) &= 1 - \phi(\mu_{k-1} - \beta X_i) \end{aligned} \quad (6)$$

TABLE 1: Characteristics and descriptive statistics of factors.

Category	factor	coding	Percentage	Mean	Std. Deviation
Fault drivers	Age	Young Age	67.5%	1.39	0.62
		Middle Age	25.5%		
		Elderly Age	7.0%		
	Gender	Male	83.8%	1.16	0.37
		Female	16.2%		
Environmental	Day of week	Working days	85.2%	1.14	0.35
		Weekend days	14.8%		
	Months	Month	Divided by month	4.14	2.93
		Late night-time	21.7%		
	Time	Daytime rush hour	44.9%	2.34	1.13
		Daytime	18.8%		
		Night-time rush hour	6.4%		
		Early night-time	8.1%		
	Weather condition	Clear	48.7%	1.71	0.78
		Fog	31.6%		
Rain		19.7%			
Work-zone information	Type of surface construction	Base	18.0%	2.97	1.43
		Asphalt	24.1%		
		Remove Asphalt	26.4%		
		Milling	6.4%		
	Number of lane closures	Concrete	25.2%	1.47	0.50
One		47.0%			
Road character	Road Character	Two	53.0%	2.34	1.26
		Straight & level	39.7%		
		Curve	14.5%		
	Road class	U-Turn	18.0%	1.32	0.47
		Straight & Grade	27.8%		
		Rural	68.1%		
Crash characteristics	Vehicle type	Urban	31.9%	0.40	0.49
		Heavy-vehicle	39.7%		
	Crash type	Other	60.3%	1.41	0.89
		Rear end	39.3%		
		Fixed Object	13.7%		
		Sideswipe	27.9%		
		Angle-side	17.2%		
		Pedestrian	1.9%		
Number of vehicle in collision	One vehicle	18.0%	0.82	0.38	
	More than one	82.0%			
Crash severity	Crash severity	No injury	24.3%		
		Injury	60.3%		
		Fatal injury	15.4%		

Here  $\Phi$  is the cumulative normal distribution function. For estimation, it can be written as

$$P(y = k) = \phi(\mu_k - \beta X_i) - \phi(\mu_{k+1} - \beta X_i) \quad (7)$$

where  $\mu_k$  depicts the lower threshold and  $\mu_{k+1}$  depicts the upper thresholds for the severity level  $k$ . Now, for all the probabilities to be positive, the threshold values must satisfy the restriction  $\mu < \dots < \mu_k < \dots < \mu_{k+1}$ . The understanding of the effect of individual estimated parameters involves the

computation of these probabilities. SPSS has been used to develop the crash model. Crash severity was considered in three levels for severity models in this paper including (i) no injury, (ii) injury, and (iii) fatal injury.

## 5. Results and Discussion

*5.1. Factor Analysis.* Factor analysis is a statistical approach used to reduce data dimensionality. In this study, factor

TABLE 2: Total variance explained.

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative %	Total	Variance (%)	Cumulative %
1	3.341	25.704	25.704	3.341	25.704	25.704	3.181	24.466	24.466
2	2.320	17.847	43.551	2.320	17.847	43.551	2.263	17.407	41.874
3	1.595	12.268	55.819	1.595	12.268	55.819	1.519	11.684	53.558
4	1.199	9.220	65.040	1.199	9.220	65.040	1.453	11.177	64.734
5	1.114	8.238	73.278	1.114	8.238	73.278	1.151	8.545	73.278
6	0.930	7.155	80.432						
7	0.688	5.293	85.725						
8	0.519	3.995	89.720						
9	0.361	2.774	92.494						
10	0.313	2.410	94.903						
11	0.303	2.331	97.235						
12	0.242	1.861	99.096						
13	0.118	0.904	100.000						

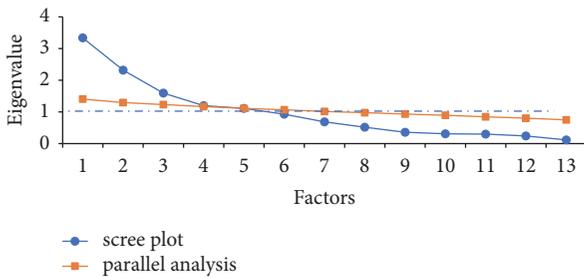


FIGURE 1: Scree plot and parallel analysis test.

analysis was conducted on the high-dimensional work-zone data to simplify the process of understanding the data. The factor analysis model was prescribed using the SPSS statistics program. To verify the suitability of the data for the analysis, the fitness test was carried out using the KMO (Kaiser-Meyer-Olkin) and Bartlett test of sphericity. The KMO test measures the adequacy of samples for each variable in the model and for the complete model. The value ranges between 0 and 1. When the KMO approaches 1, it implies a high correlation between variables, and the data is suitable for factor analysis. In this paper, the KMO value of 0.67 obtained suggests that the data is well suited for factor analysis [37, 38]. In the case of Bartlett test of sphericity, the Chi-square value of 1786.156, with 78 degrees of freedom, was found to be significant at 0.000. The result here demonstrates a good correlation between variables and proved to be suitable for factor analysis. To ensure reliable results, we apply three tests, i.e., the eigenvalue, scree plot, and parallel analysis test to identify the number of factors that reasonably explain greater percentage of the data variability. The eigenvalue (EV) represents the proportion of variance accounted for by each factor. In Table 2, from the initial eigenvalue, the percentage of variance explained by the first principal factor was 24.466% and the percentage explained by

TABLE 3: Parallel analysis results.

Roots	Means	Percentile
1.0	1.34320	1.40688
2.0	1.25312	1.29635
3.0	1.18425	1.23144
4.0	1.12981	1.17157
5.0	1.08060	1.11330
6.0	1.03771	1.06735
7.0	0.98988	1.0182
8.0	0.94894	0.9767
9.0	0.90320	0.93548
10.0	0.85773	0.89285
11.0	0.80870	0.84597
12.0	0.76053	0.80232
13.0	0.70235	0.75033

the second, third, fourth, and fifth factors is 17.407%, 11.684%, 11.177%, and 8.543%, respectively. The cumulative percentage variance explained by the five factors was 73.278%. In broad terms, considering the explanatory power of a factor, we retain the factors with eigenvalue more than 1 as a common factor, consistent with the first five factors.

The number of common factors also can be known by parallel analysis test and scree plot. Figure 1 shows a scree plot and parallel analysis test consisting of 13 factors. From the fifth factor, the eigenvalue of each factor declined more slowly. Table 3 illustrates the results of parallel analysis test for each factor. According to the results of the test methods, we should select the first five factors as common factors. It is significant to state that one of the motives for conducting factor analysis in this study is to decrease the large number of variables that describe a complex phenomenon such as work-zone crash severity analysis to limited interpretable

TABLE 4: Rotated component matrix.

Factor	Component				
	F1	F2	F3	F4	F5
Number of lane closures	<b>.901</b>	-.046	-.078	.058	-.007
Type of surface construction	<b>.848</b>	.046	-.027	-.014	.064
Road Character	<b>.830</b>	-.029	-.014	-.020	-.053
Road class	<b>.813</b>	.052	-.055	-.051	.047
Day of week	-.084	<b>.887</b>	.003	.005	.043
Month	.092	<b>.852</b>	-.164	.028	-.013
Time	.017	<b>.843</b>	.066	-.122	.064
Crash type	-.172	.000	<b>.869</b>	-.127	.065
Number of the vehicles	.459	-.031	<b>-.603</b>	.280	-.157
Vehicle type	.162	-.094	<b>.593</b>	.178	-.156
Gender	-.106	.026	.015	<b>.825</b>	-.219
Age	.081	-.125	-.083	<b>.790</b>	.344
Weather condition	.027	.084	-.010	.017	<b>.936</b>

TABLE 5: Name of the common factors.

Influential factors	New name	Variables
Factor 1	Advanced work-zone information	Number of lane closures, type of surface construction, road character, and road class
Factor 2	Visibility and traffic condition	Day of week, month, and time
Factor 3	Crash characteristic	Crash type, number of the vehicles, and vehicle type
Factor 4	Driver skills	Gender and age
Factor 5	Weather condition	Weather condition

latent variables known as a factor. In essence, this section seeks to identify a smaller number of interpretable factors that explain the maximum amount of variability in the data. The principal component method was used to extract the factor loading matrix. The varimax orthogonal method has also been used to diffuse the factors that are common, by factor rotation. The loadings of the rotated factor loading matrix further differed from each other. High loadings were dispersed close to the matrix diagonal, which suggested that each common factor is mainly associated with a few factors. This study adopts a strategy which is named the common factor, based on the communal idea of influential variables taken into account in the common factors. In this way, we can better comprehend the function implication of each common factor. The five factors with high loadings as shown in Table 4 are (number of lane closures = .901, type of surface construction = .830, road character = .830, and road class = .813) related to work-zone information being a feasible and functional countermeasure for reducing the severity of crashes occurring in work-zones. Similarly, the four factors with high loadings (day of week = .887, month = .852, and time = .843) on factor two reveal information about visibility and traffic condition. Factor three reveals information about the crash characteristics (crash type = .869, number of the vehicles = -.603, and vehicle type = .593). In the case of factor four, the variables with high loading are gender and age, all related to driver skills and ability to operate safely

under different work-zone configurations. Finally, there is the weather condition factor. The names assigned to the five common factors are shown in Table 5.

Based on the regression method factor score matrix was extracted. Each powerful factor had scored on every common factor. Through normalizing of the variance percentage of the five common factors, the weights of factors were calculated. The normalized weight  $A_i$  is estimated using the data of rotated loading quadratic sum. The absolute value  $A_k$  ( $k$  is the number of influential factors) of the influential factor score of each common factor was extracted in Table 6, and then normalization was proposed. The score  $A_k$  after normalization represents the weight of the influential factor in relevant common factor. Then the weight  $W_k$  of the influential factors of the work-zone crashes was deduced. The weights of main influential factor are shown in Table 7. Figure 2 represents the weight of 13 influential factors  $W_k$ .

$$W_k = A_i \cdot A_k \quad (8)$$

The results showed that the weather condition factor is the most important factor that explains high amount of the variability in the original 13 variables, followed by number of lane closures, type of surface construction, road character, day of week, and so forth. The first five influential factors had a noticeable effect on work-zone crashes because their individual weights are significant (i.e.,  $\geq 8\%$ ).

TABLE 6: Factor score matrix.

Factor	Component				
	F1	F2	F3	F4	F5
Age	.005	-.029	.027	<b>.541</b>	.305
Gender	-.043	.077	.099	<b>.600</b>	-.217
Time	.019	<b>.373</b>	.072	-.027	.010
Day of week	-.023	<b>.399</b>	.032	.061	-.012
Month	.017	<b>.381</b>	-.065	.055	-.059
Type of surface construction	<b>.279</b>	.019	.072	-.022	.053
Weather condition	.004	-.015	-.018	-.003	<b>.845</b>
Number of lane closures	<b>.290</b>	-.015	.046	.018	-.007
Road Character	<b>.274</b>	-.007	.077	-.027	-.050
Road class	<b>.265</b>	.019	.044	-.051	.039
Vehicle type	.116	.006	<b>.461</b>	.190	-.154
Crash type	.038	.026	<b>.587</b>	.009	.044
Number of the vehicles	.084	-.012	<b>-.348</b>	.127	-.135

TABLE 7: The weight of each influential factor.

Common factor	Weight of the common factor $A_i$	Factor	Absolute value	The weight of factor on corresponding common factor (normalization) $A_k$	Weight of factor $WK$
Advanced work-zone information	0.333	Number of lane closures	0.29	0.262	0.087
		Type of surface construction	0.279	0.252	0.084
		Road Character	0.274	0.247	0.082
		Road class	0.265	0.239	0.080
Visibility and traffic condition	0.238	Day of week	0.399	0.346	0.082
		Month	0.381	0.330	0.078
		Time	0.373	0.324	0.077
Crash characteristic	0.162	Crash type	0.587	0.420	0.068
		Number of the vehicles	0.348	0.249	0.040
		Vehicle type	0.461	0.330	0.053
Driver skills	0.152	Gender	0.6	0.526	0.080
		Age	0.541	0.474	0.072
Weather condition	0.116	Weather condition	0.845	1.000	0.116
<b>Total</b>	<b>1</b>			<b>5</b>	<b>1</b>

The results in this section are consistent with findings in the literature. The impact of weather condition on work-zone crash severity is consistent in the research by [39]. Regarding traffic safety at freeway work-zone, Ozturk et al. [40] and WuBiao et al. [41] concluded that number of lane closures has a significant relationship. Road character and day of week were found to be significantly associated with work-zone severity [9, 10, 37, 39]; these conclusions were consistent with the findings in this paper.

*5.2. Estimations of Ordered Probit Model.* The ordered probit model was specified using SPSS. The analysis consisted of 345 observations of work-zone crashes on 13 factors. The factor

analysis identified five factors, which were used as inputs for the ordered probit model. The factors were tested for statistical significance ( $p$  value  $< 0.05$ ) to find out the factors that significantly influence injury severity crashes at highway construction work-zones in Egypt. The results of the ordered probit model are presented in Table 8. Four factors were identified as significant factors associated with crash injury severity in work-zones.

As estimated in the ordered probit model, the crash characteristics factor has a better safety performance in terms of reducing the probability of the average injury risk, as compared with advanced information factor, driver skills factor, and weather condition factor. In other words, crashes

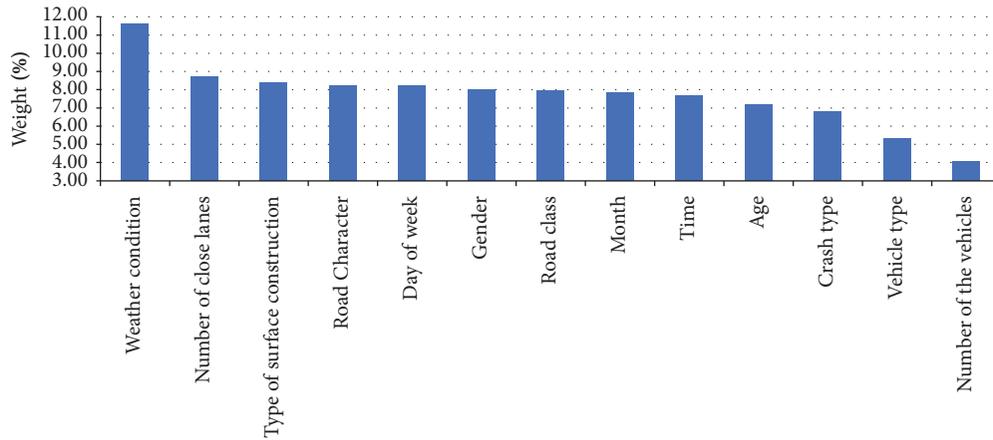


FIGURE 2: Influential factors' weight ranking.

TABLE 8: Ordered probit model estimation results.

		Estimate	Std. Error	Wald	DF	Sig ( $\alpha$ )	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	No injury	-.738	.076	93.541	1	.000	-.887	-.588
	Injury	1.086	.085	161.719	1	.000	.919	1.254
Advanced work-zone information factor		.161	.062	6.711	1	.010	.039	.284
Visibility and traffic factor		.108	.062	3.057	1	.080	-.013	.230
Crash characteristic factor		-.223	.064	12.275	1	.000	-.349	-.098
Driver skills factor		-.140	.063	4.990	1	.025	-.263	-.017
Weather condition factor		.132	.062	4.478	1	.034	.010	.253

that occurred under the influence of crash characteristic factor are more likely to have a nonfatal injury than under other factors. Based on the result, the visibility and traffic factor does not have a significant impact on injury severity of work-zone traffic crashes. The coefficient of advanced work-zone information factor is found to be positive and thus points to the fact that inadequate advanced information about construction work-zones, on Egyptian highways, tends to increase the probability of fatal injury crashes in work-zones. In this regard, [42] demonstrated that road traffic crashes occur for one of three key factors. The first reason, related to information flow, is a perceptual error. This occurs, for instance, when work-zone conditions are not perceived early enough to allow for sufficient driver perception-reaction time.

The increase in driver skills factor will decrease the probability of fatal injury in work-zone crashes. One possible reason may be the lack of sufficient driving skills for safe maneuver in the event of unfavourable traffic and roads. Several studies have been conducted to investigate the factors, relating to driver skills, which are influential in predicting fatal injury crashes. A study by [43] found that young drivers are overrepresented in road crashes and fatalities and that one approach to improving their safety is the provision of advanced training. Beanland et al. [44] reviewed the literature on the efficacy of advanced driver training with the aim

being to assess its effectiveness as a means of reducing young driver crash-involvement. Their report indicates that various forms of prelicense and postlicense training have been found to be effective, for skill acquisition and enhanced driving performance. The results revealed that crashes that occurred under the influence of weather condition factor have higher probability of fatal crashes. This finding is consistent with previous studies [15, 39] and in contrast with the previous study [13].

The probability to be involved in no injury and injury crash rather than fatal crash will increase significantly if the crash occurred under the influence of crash characteristic factor. That may be explained by the fact that crashes involving heavy vehicles often result in multiple-vehicle crashes as opposed to when crashes occur without heavy vehicles, because heavy vehicles have reduced breaking abilities. This is explained by the excessive weight carried on trucks in Egypt; more than 96% of Egypt's goods are transported on trucks [45].

Hence, the influence of the factors on the risk of the crash can be ranked by comparing the size of the estimated coefficients of the factors. In that regard, it can be concluded that advanced information factor apparently has the highest and greatest impact on injury level for a crash occurring at a work-zone (with  $\beta = .161$ ), whereas weather condition factor appears to have a relatively low risk ( $\beta = .132$ ). In

terms of magnitude, for example, crashes that occurred under the influence of advanced information factor are 1.22 times crashes that occurred under the influence of factor weather condition, assuming all factors remain being equal.

## 6. Conclusion and Recommendations

In this paper, a hybrid approach, which combines an ordered probit model and factor analysis method, was developed to carry out a comprehensive analysis of the injury severity of crashes in highway construction work-zones in Egypt. The results of the factor analysis revealed the important and main factors that are influential in determining crash severity were weather condition, number of lane closures, type of surface construction, road character, day of the week, and so forth. The advanced work-zone information factor, visibility and traffic factor, crash characteristic factor, driver skills factor, and weather condition factor are common factors. The percentage variance explained by the five factors is 73.278%. We applied the results of the factor analysis method to calibrate the ordered probit model to examine the influence of these factors in predicting injury severity of work-zone crashes. The model estimation results showed that four factors (advanced work-zone information factor, crash characteristic factor, driver skills factor, and weather condition factor) are significant risk factors associated with work-zone crashes. In addition, the results showed that the weather condition factor had significantly great influence on severe work-zone crashes in Egypt.

Based on the results obtained in this study, four factors which had a massive influence on severe injury work-zone crashes were found in this study. This means that the proportion of 55.87% variance in work-zone risk could be explained by four factors. Thus, this paper offers the following suggestions that can help to reduce the frequency and severity of work-zone crashes, for improved traffic safety in highway construction work-zones in Egypt.

Driver skills training: programmes designed to offer regular skills training to prelicensed and postlicensed drivers are required to reduce the incidence of bad driving behavior, for improved road safety. Besides, existing countermeasures that work well in other countries can be adapted to Egypt, such as graduated driving licensing programmes [46–48] and additional precincts to minimize exposure of novice drivers to work-zone crashes [49, 50].

Provision of advanced information: the fact that most of the crashes that occurred were associated with human errors infers that advanced knowledge of road and traffic conditions will assist drivers to prepare and take action to avoid crashes adequately. In this regard, we recommend the development of effective education programmes to educate the general public as well, who have to travel through work-zones. In addition, ITS technologies such as Dynamic Message Signs (DMS), at a reasonable distance ahead of the work-zone, may be an effective means of compensating for adverse design, human factors, and roadway conditions.

This paper has some limitations, which may affect the result and interpretations; some information (seat belt, alcohol, etc.) was not taken into account, due to the lack of such

information in the database. It is recommended that all of this information be collected in the work-zone crash database and used for model calibration in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors thank the Ministry of Commerce of China (MOFCOM) for funding of Ph.D. doctoral program at Southwest Jiaotong University and Dr. Essam Radwan from University of Central Florida for valuable comments and suggestions by that have helped improve this article.

## References

- [1] World Health Organization, *Global Status Report on Road Safety 2015, WHO Library Cataloguing-in-Publication Data Global*, 2015.
- [2] M. Abdel-Aty, "Analysis of driver injury severity levels at multiple locations using ordered probit models," *Journal of Safety Research*, vol. 34, no. 5, pp. 597–603, 2003.
- [3] A. Khattak, "Injury Severity in Multivehicle Rear-End Crashes," *Transportation Research Record*, vol. 1746, pp. 59–68, 2001.
- [4] K. M. Kockelman and Y.-J. Kweon, "Driver injury severity: an application of ordered probit models," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 313–321, 2002.
- [5] S. Dissanayake and J. Lu, "Analysis of severity of young driver crashes: Sequential binary logistic regression modeling," *Transportation Research Record*, no. 1784, pp. 108–114, 2002.
- [6] M. A. Abdel-Aty, C. L. Chen, and J. R. Schott, "An assessment of the effect of driver age on traffic accident involvement using log-linear models," *Accident Analysis & Prevention*, vol. 30, no. 6, pp. 851–861, 1998.
- [7] E. Donnell, D. Harwood, K. Bauer, J. Mason, and M. Pietrucha, "Cross-Median Collisions on Pennsylvania Interstates and Expressways," *Transportation Research Record*, vol. 1784, pp. 91–99, 2002.
- [8] E. Donnell and J. Mason, "Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression," *Transportation Research Record*, vol. 1897, pp. 55–63, 2004.
- [9] M. Osman, R. Paleti, and S. Mishra, "Analysis of passenger-car crash injury severity in different work zone configurations," *Accident Analysis & Prevention*, vol. 111, pp. 161–172, 2018.
- [10] N. N. Sze and Z. Song, "Factors contributing to injury severity in work zone related crashes in New Zealand," *International Journal of Sustainable Transportation*, pp. 1–7, 2018.
- [11] Y. Li and Y. Bai, "Development of crash-severity-index models for the measurement of work zone risk levels," *Accident Analysis & Prevention*, vol. 40, no. 5, pp. 1724–1731, 2008.
- [12] A. Theofilatos, A. Ziakopoulos, E. Papadimitriou, G. Yannis, and K. Diamandouros, "Meta-analysis of the effect of road work

- zones on crash occurrence," *Accident Analysis & Prevention*, vol. 108, pp. 1–8, 2017.
- [13] H. Yang, K. Ozbay, O. Ozturk, and K. Xie, "Work Zone Safety Analysis and Modeling: A State-of-the-Art Review," *Traffic Injury Prevention*, vol. 16, no. 4, pp. 387–396, 2015.
- [14] H. Yang, K. Ozbay, O. Ozturk, and M. Yildirimoglu, "Modeling work zone crash frequency by quantifying measurement errors in work zone length," *Accident Analysis & Prevention*, vol. 55, pp. 192–201, 2013.
- [15] O. Ozturk, K. Ozbay, and H. Yang, "Estimating the Impact of Work Zones on Highway Safety," *TRB 93rd Annual Meeting - Transportation Research Board*, vol. 116, pp. 1–16, 2014.
- [16] L.-Y. Chang and F. Mannering, "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents," *Accident Analysis & Prevention*, vol. 31, no. 5, pp. 579–592, 1999.
- [17] K. Krull, A. Khattak, and F. Council, "Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes," *Transportation Research Record*, vol. 1717, pp. 46–54, 2000.
- [18] A. Khattak and F. Targa, "Injury Severity and Total Harm in Truck-Involved Work Zone Crashes," *Transportation Research Record*, vol. 1877, pp. 106–116, 2004.
- [19] A. J. Khattak, R. J. Schneider, and F. Targa, "Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions," *Transportation Research Record: Journal of the Transportation*, vol. 40, 2003.
- [20] A. J. Khattak, M. D. Pawlovich, R. R. Souleyrette, and S. L. Hallmark, "Factors related to more severe older driver traffic crash injuries," *Journal of Transportation Engineering*, vol. 128, no. 3, pp. 243–249, 2002.
- [21] M. Abdel-Aty and J. Keller, "Exploring the overall and specific crash severity levels at signalized intersections," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 417–425, 2005.
- [22] R. Garrido, A. Bastos, A. de Almeida, and J. P. Elvas, "Prediction of Road Accident Severity Using the Ordered Probit Model," *Transportation Research Procedia*, vol. 3, pp. 214–223, 2014.
- [23] Y. Qi, R. Srinivasan, H. Teng, and R. Baker, "Analysis of the Frequency and Severity of Rear-End Crashes in Work Zones," *Traffic Injury Prevention*, vol. 14, no. 1, pp. 61–72, 2013.
- [24] V. Katta, *Development of crash severity model for predicting risk factors in work zones for Ohio*, 2013.
- [25] J. Weng and Q. Meng, "Effects of environment, vehicle and driver characteristics on risky driving behavior at work zones," *Safety Science*, vol. 50, no. 4, pp. 1034–1042, 2012.
- [26] H. Xu, "Study on road traffic accident data analyzing and mining," *J Chin People Public Sec Univ*, vol. 4, pp. 69–73, 2008.
- [27] P. Sun, R. Song, and H. Wang, "Prevention and analysis of the traffic accidents," *Safe Environment Engineering*, vol. 14, pp. 97–100, 2007.
- [28] W. Cheng, *Research Based on Model And Method of Urban Traffic Accidents And Traffic Conflict Technique*, China Jilin Univ, Jilin, China, 2004.
- [29] T. Chen, C. Zhang, and L. Xu, "Factor analysis of fatal road traffic crashes with massive casualties in China," *Advances in Mechanical Engineering*, vol. 8, no. 4, p. 168781401664271, 2016.
- [30] Q. Yuan, X. Li, C. Wang, Y. Li, and Y. Gao, "Cluster and factor analysis on data of fatal traffic crashes in China," in *Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp. 211–224, Banff, AB, Canada, August 2017.
- [31] D. Harper, J. Kim, and C. W. Mueller, "Introduction to Factor Analysis: What It Is and How to Do It.," *Contemporary Sociology*, vol. 9, no. 4, p. 562, 1980.
- [32] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, NY, USA, 1984.
- [33] R. D. McKelvey and W. Zavoina, "A statistical model for the analysis of ordinal level dependent variables," *The Journal of Mathematical Sociology*, vol. 4, no. 1, pp. 103–120, 1975.
- [34] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall, Boca Raton, Fla, USA, 2003.
- [35] C. J. O'Donnell and D. H. Connor, "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice," *Accident Analysis & Prevention*, vol. 28, no. 6, pp. 739–753, 1996.
- [36] F. Ye and D. Lord, "Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models," *Analytic Methods in Accident Research*, vol. 1, pp. 72–85, 2014.
- [37] S. A. Mulaik, *Foundations of factor analysis*, Statistics in the Social and Behavioral Sciences Series, CRC Press, Boca Raton, FL, USA, 2nd edition, 2010.
- [38] B. G. Tabachnick, "Using Multivariate Statistics," *Bost Allyn Bacon Bamberger*, vol. 31, p. 1, 2001.
- [39] R. Harb, E. Radwan, X. Yan, A. Pande, and M. Abdel-Aty, "Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression," *Journal of Transportation Engineering*, vol. 134, no. 5, pp. 203–214, 2008.
- [40] O. Ozturk, K. Ozbay, H. Yang, and B. Bartin, "Crash Frequency Modeling for Highway Construction Zones," 2013.
- [41] B. Wu, H. Xu, and W. Zhang, "Identifying the Cause and Effect Factors of Traffic Safety at Freeway Work Zone Based on DEMATEL Model," in *Proceedings of the Second International Conference on Transportation Engineering*, pp. 2183–2188, Southwest Jiaotong University, Chengdu, China.
- [42] M. Green and J. Senders, *Human error in road accidents, visual expert*, Canada, CRC Press, 2018.
- [43] S. Gargett, L. B. Connelly, and S. Nghiem, "Are we there yet? Australian road safety targets and road traffic crash fatalities," *BMC Public Health*, vol. 11, 2011.
- [44] V. Beanland, N. Goode, P. M. Salmon, and M. G. Lenné, "Is there a case for driver training? A review of the efficacy of pre- and post-licence driver training," *Safety Science*, vol. 51, no. 1, pp. 127–137, 2013.
- [45] A. F. Elshamly, R. A. El-Hakim, H. A. Afify et al., "Factors Affecting Accidents Risks among Truck Drivers In Egypt," *Matec Web of Conferences*, vol. 124, p. 04009, 2017.
- [46] L. T. Lam, "Factors associated with young drivers' car crash injury: Comparisons among learner, provisional, and full licensees," *Accident Analysis & Prevention*, vol. 35, no. 6, pp. 913–920, 2003.
- [47] A. T. McCartt and E. R. Teoh, "Tracking progress in teenage driver crash risk in the United States since the advent of graduated driver licensing programs," *Journal of Safety Research*, vol. 53, pp. 1–9, 2015.
- [48] A. F. Williams, "Teenage drivers: Patterns of risk," *Journal of Safety Research*, vol. 34, no. 1, pp. 5–15, 2003.
- [49] A. E. Curry, M. R. Pfeiffer, D. R. Durbin, and M. R. Elliott, "Young driver crash rates by licensing age, driving experience, and license phase," *Accident Analysis & Prevention*, vol. 80, pp. 243–250, 2015.
- [50] A. T. McCartt, D. R. Mayhew, K. A. Braitman, S. A. Ferguson, and H. M. Simpson, "Effects of age and experience on young driver crashes: Review of recent literature," *Traffic Injury Prevention*, vol. 10, no. 3, pp. 209–219, 2009.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

