Selecting Qualitative Features of Driver Behavior via Pareto Analysis

Kawtar Zinebi  
Mohammed V University in Rabat. EMI - SIWEB Team. Rabat, Morocco  
Email: kawtarzinebi@gmail.com

Nissrine Souissi  
Mohammed V University in Rabat. EMI - SIWEB Team. Rabat, Morocco  
Email: souissi@enim.ac.ma

Kawtar Tikito  
École Nationale Supérieure des Mines de Rabat. Computer Science Department. Rabat, Morocco  
Email: tikito@enim.ac.ma

Received: 21 July 2018; Accepted: 05 September 2018; Published: 08 October 2018

Abstract—Driver behavior is the main cause of road crashes; it is the key element that insures a better understanding and improves predictions of car accidents. The main goal of our study is to determine the set of driver behavior features that are the most encountered in literature; we were based on behavioral questionnaires as a source for these features. We selected the questionnaires that are most cited in literature and therefore proved their efficiency through many studies they were employed in. Then we extracted the features considered in their items and classified them by rate of appearance according to the Pareto & ABC principle. In the second part of our study we collaborated with the National Committee for Circulation Accident Prevention (CNPAC) of the Ministry of Transportation of Morocco in order to compare the findings we gathered from literature with the researches they administer. We prepared a questionnaire that contains the final set of features and we transmitted it to experts working in the road safety field to rate it according to their knowledge and experience. Data analysis showed significant differences in some features, which demonstrates the gap between theoretical results and field research.

Index Terms—Driver behavior features, behavioral questionnaires, Pareto & ABC analysis, DBQ, DAX, MDSI, DSQ, DBI.

I. INTRODUCTION

Vehicle related causalities are one of the most deadly causes worldwide. According to the World Health Organization, traffic accidents caused about 1.25 million deaths in 2015, 75% are male victims and 49% are among pedestrians, cyclists and motorists. The same report shows that car crashes are the first cause of deaths among population aged between 15-29 years.

Many aspects are taken into account in defining causes of crash frequency and injuries severity, they can be classified into infrastructure, vehicle defects and human faults [1]. The first category consists of road traffic characteristics such as type of road and its conditions, weather and traffic flow, the second one includes conditions of the vehicle, and the last one involves around driver’s characteristics, like driving experience, fatigue, physiological and psychological state, personality traits and attitudes.

Driver behavior proves to be the key element behind car crashes involvement [2], as data show that 70–80% of road accidents were completely the fault of drivers, and this rate is well known to be similar in the majority of countries in the world, which is why vigorous efforts are under way today to fully apprehend and model driver behavior.

The description of such a system is not an obvious task, given the complexity and the diversity of activities and functions performed while driving. Many studies address driver’s poor perception as the main feature for accident involvement [2,3,4,5], others reported about risk taking to have a major impact on road crashes especially among young and male drivers [6,7,8].

The main difficulty that researchers encounter is the absence of a unified definition and a standard method to model driver behavior. We realized that literature lacks a common model; it needs to acquire a universal model given its importance for safety organizations, industrial manufacturers and researchers all around the globe. Research on this topic had increased exponentially during the last decade considering the technological evolution of electronic devices and communication protocols, but still there is a deficiency in terms of standardized models that define driver behavior as a whole system, and not just portions of it. Therefore we decide to address such behavior by adopting a generic approach that doesn’t
focus on one or two features, but instead it considers the entirety of its elements in order to represent the multi-dimensional nature of this phenomenon.

This is why our main objective in this paper is to collect the set of features that control driver behavior as a first step into defining a generic model for it. The main resources we consider for these features are behavioral questionnaires; our methodology is based on these questionnaires as an input in order to extract major features that need to be taken into account in a behavioral model.

Our paper is organized as following: in the second section, we present our research method in details and explain the different steps we followed, next in the third section, we address each questionnaire and classify the features we extracted in them, and finally in the fourth section we conduct a comparative study with the CNPAC of Morocco in order to compare theoretical results from literature with their field projects.

II. RESEARCH METHOD

The main reason we chose to work with behavioral questionnaires is the diversity of their items and the large amount of information they contain. They represent many situations that drivers are likely to encounter on the roads, that others collecting methods might miss. They also summarize the circumstances of these situations; their items express explicitly the causes of such behaviors and indicate the intents of the driver behind them. They also question about the driver’s state of mind, feelings, personal motivations and attitudes toward driving, which is a benefit for our research since our main goal is to collect as many features as possible, from quantitative measures (speed, acceleration etc.) to qualitative features (personality, state of mind, fatigue etc.)

In this paper we are interested in analyzing the questionnaires items, their design and the information they include, we will not address the results of such measurements since this topic brings many controversy about reliability and bias.

Step 1: Selecting driver behavior questionnaires

In this step, we represent the questionnaires we selected for our study. The main selection criteria are the number of citation and the diversity of themes. In table 1, we gathered the set of the most cited questionnaires in literature. For this purpose, we used the research motors ScholarGoogle, IEEXplore and ScienceDirect.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All citations</td>
<td>1132</td>
<td>252</td>
<td>201</td>
<td>46</td>
</tr>
<tr>
<td>All citations</td>
<td>206</td>
<td>240</td>
<td>476</td>
<td>96</td>
</tr>
</tbody>
</table>

This table shows that literature includes eight behavioral questionnaires as the most employed in traffic psychology studies. We chose to eliminate three questionnaires from this list for the main reason of avoiding redundancy; the items presented in the Reckless Driving Habit Scale, the DVQ and the DAS are all reproduced in the others questionnaires.

Finally we ended up with five questionnaires indicated in the list below.

- **Driver Behavior Questionnaire (DBQ) [9]:** is one of the most widely used method to assess self-reported driving behaviors [17,18,19], its main objective is to help understanding the relation between violations, crash involvement and personal driving habits.
- **Driving Anger Expression Inventory (DAX) [10]:** is a measurement developed to evaluate how drivers express their anger on the road, its items describe different situations driven by irritation or frustration form traffic or other drivers behavior.
- **Multidimensional Driving Style Inventory (MDSI) [11]:** it targets eight main styles: “dissociative, anxious, risky, angry, high-velocity, distress reduction, patient, and Carefulness”. It was designed to assess driver’s habitual behaviors and driving-specific features that can explain accident involvement and traffic law violations, it takes into account general socio-demographic and personality features
- **Driving Style Questionnaire (DSQ) [14]:** it contains 15 items that target what the authors called driving dimensions: speed, calmness, social resistance, focus, planning, and deviance.
- **Driving Behavior Inventory (DBI) [15]:** It is based on four elements which are driving aggression, dislike of driving, tension and frustration connected with successful or unsuccessful overtaking, irritation when overtaken and heightened alertness and concentration. The 35 items of the questionnaire highlights driver stress reactions to different situations.

Step 2: Extracting driver behavior features

After selecting the questionnaires, the next step is to extract behavior features from them. This operation depends on the items of each questionnaire. As an example, the first item in the DBQ is “Attempt to drive away from traffic lights in third gear”; the situation presented in this item contains one behavioral feature which is “Inattention”. Another item from the DBQ is “Wake up to realize that you have no clear recollection of the road along which you have just travelled”, it mainly refers to “Day dreaming”. More details of this phase are presented in the third section of our paper.

Step 3: Pareto & ABC analysis

Pareto analysis is a technique that helps identifying the top most important elements in terms of appearance frequency. It is widely used by industrial management in order to diagnose failures in any kind of systems, its
principle is based on classifying causes of problems and going to take advantage of the ordering aspect on the Pareto & ABC analysis to determine the predominant features taken into account in each questionnaire. This operation will help us acknowledge the most essential parameters that conduct driver behavior.

**Step 4: Comparison study with the CNPAC**

This step is the final phase of our study, it consists of validating the previous findings from step3. We need to compare our theoretical results with an accredited organization so we can deduce the gap between field studies and theoretical research. For this, we collaborated with the National Committee for Circulation Accident Prevention (CNPAC) of the Ministry of Transportation of Morocco in order to evaluate our findings from literature with their statistical studies.

**III. DRIVER BEHAVIOR QUESTIONNAIRES & FEATURES EXTRACTION**

In this section, we represent the details of step3 and the results of Pareto & ABC analysis for all the questionnaires. Before that, we give a general presentation of the five questionnaires we analyze.

**A. Questionnaires Presentation**

As mentioned in step1, our analysis is based on five questionnaires which we selected based on many criteria, namely their design and dimensions. We made sure the questionnaires addresses different aspects of driver behavior so we avoid repetitiveness. The main concepts we managed to frame are Risk taking, Anxiety, Anger, Mistakes and Inattention. These four concepts are a summary of the features presented in the questionnaires items. We created this four-dimension representation in order to demonstrate that these concepts are independently coherent and perpendicular to each other, which implies that they can be seen as independent and distinctive aspects of driver behavior.

On the other hand, even though the questionnaires’ themes are disparate, they have some intersections points as shown in table2.

As an example for the representation of these aspects in the questionnaires items, stress is illustrated in the MDSI by items such as “feel nervous while driving” and by “I feel anxious when overtaken at a junction” from the DBI. Revenge is represented in the DAX by “I do to other drivers what they did to me” and “Angered by another driver's behavior, you give chase with the intention of giving him/her a piece of your mind” from the DBQ.

**B. Pareto & ABC Analysis**

In this section, we apply Pareto & ABC analysis for each questionnaire and we represent their results.

1) DBQ:

The results of Pareto & ABC analysis for the DBQ are presented in figure 1.

<table>
<thead>
<tr>
<th>DBQ</th>
<th>MDSI</th>
<th>DAX</th>
<th>DBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDSI</td>
<td>Slips &amp; Mistakes &amp; Lapses &amp; Violations</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>DAX</td>
<td>Use of vehicle for aggression</td>
<td>Angry driving &amp; Careful driving</td>
<td>*</td>
</tr>
<tr>
<td>DBI</td>
<td>———</td>
<td>Anxious driving &amp; angry driving</td>
<td>Adaptive expression &amp; Use of vehicle for aggression</td>
</tr>
<tr>
<td>DSQ</td>
<td>Deviance &amp; Planning &amp; Speed</td>
<td>Speed &amp; calmness &amp; Deviance &amp; Focus</td>
<td>Speed</td>
</tr>
</tbody>
</table>

Results of the DBQ show different percentages of various elements; Inattentiveness, Overtaking, Bad perception, Speed, Turning, Absent-minded. Not respecting signs and Aggressive braking have the most rate of appearance as they belong in the class A, they show up in items like “Distracted or preoccupied, realize belatedly that the vehicle ahead has slowed, and have to slam on the brakes to avoid a collision”, “Cut the corner on a right-hand turn and have to swerve violently to avoid an oncoming vehicle” and “Disregard red lights when driving late at night along empty roads”.

We remind that this classification does not mean that features with least rate of appearance are least significant, we don’t judge the features by their importance or impact on accident involvement, we order them according to their appearance frequency in the questionnaire items.
Selecting Qualitative Features of Driver Behavior via Pareto Analysis

2) DAX:

We applied the same Pareto principle as the previous questionnaire. Figure 2 illustrates the results for the DAX.

According to these results, the elements that are most present in the DAX are Aggressive expression and Carefulness. These findings are completely predictable since the questionnaire is orientated toward hostile thinking and assertive behaviors. The aggressive expressions are a combination of many acts like glaring at other drivers, shaking head as an expression of anger, yelling things like “where did you get your license?”, and getting out of the car to tell others to get out of the way, while the careful actions are presented by items such as “I pay even closer attention to being a safe driver” and “I pay even closer attention to other’s driving to avoid accidents”.

3) MDSI:

The Multidimensional Driving Style Inventory is originally designed by [11], it has demonstrated its sufficiency through many studies aiming to evaluate the correspondence between personality traits and driving habits [20,21,22] and it has been translated into many languages [23,24]. The Pareto analysis for its items is presented in figure 3.
The figure shows that the features most included are Anxiousness, Sensation seeking, Anger, Use of vehicle, Carefulness, Patience, Bad perception, Inattention and Route planning. We can make the first remark about the dominant themes of the MDSI, which are emotion-driven actions. These actions are presented by items such as “blow my horn or “flash” the car in front as a way of expressing frustrations”, “feel nervous while driving” and “nearly hit something due to misjudging my gap in a parking lot”.

4) DSQ:
The Pareto analysis for the DSQ is represented in figure 4.

According to these results, the features that appear the most in the DSQ are Carefulness, Speed and Anxiousness. These features are directly related to the objective of the questionnaire, which is modeling decision making of drivers based on their driving style. These features are present in items such as “Do you exceed the speed limit in built up areas”, “Is your driving affected by pressure from other motorists?” and “Do you find it easy to ignore distractions while driving”.

5) DBI:
The Pareto analysis for the DBI is represented in figure 5.
Results show that Anxiousness is the most cited feature, which is entirely predictable since the theme of the DBI is adjusted on driver stress. Overtaking also proves to be much referred to in items such as “I feel angry when overtaken at a junction”, “I do not feel indifferent when overtaking another vehicle”, “I feel satisfied when overtaking another vehicle” and “I feel tense when overtaking another vehicle”.

C. Classification and Synthesis

In this section, we gather all the features that we extracted from all the questionnaires above, and we calculate the number of times they occurred as a Class A, Class B, Class C. The results are presented in table 3.

<table>
<thead>
<tr>
<th>Features</th>
<th>Class A</th>
<th>Class B</th>
<th>Class C</th>
<th>Not mentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiousness</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Carefulness</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Anger</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bad perception</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Inattention</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Overtaking</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Speed</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Absent-minded</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Aggressive braking</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Aggressive expression</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Not respecting signs</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Patience</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Route planning</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Sensation seeking</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Turning</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Use of vehicle</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Distraction</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Forgetfulness</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Lane changing</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Revenge</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Slips</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Tailgating</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

All of these driving-related features are crucial in predicting driving styles, as they figure in the most cited driving behavior questionnaires. We remind that this classification does not reflect the importance nor the influence of these features on the gravity of accidents, but this categorization illustrates their consideration rate in self-report measures and behavioral studies in traffic safety field.

IV. COMPARATIVE ANALYSIS: CNPAC CASE STUDY

A. Participants and Procedure

The final step of our analysis is to compare theoretical results collected from scientific literature with a real life case of study. To do so, we collaborated with the National Committee for Circulation Accidents Prevention (CNPAC) of the Ministry of Transport in Morocco. We prepared a questionnaire that contains the final set of features mentioned above and we transmitted it to experts working in the road safety field to answer it according to their knowledge and experience.

The questionnaire is a 4 point Likert scale, we chose this scale so we keep coherence between the testing data we get from this questionnaire and the previous theoretical results we had from the literature review. We associated the Class A, Class B, Class C and Not mentioned respectively with the Likert scale choices “4” (almost always), “3” (sometimes), “2” (every once in a while) and “1” (rarely) as shown in table 4.

<table>
<thead>
<tr>
<th>Theoretical scale</th>
<th>Not mentioned</th>
<th>Class C</th>
<th>Class B</th>
<th>Class A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing data scale</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The comparison process begins by collecting data from answers of the questionnaire; for each feature we calculated the number of person who checked choices “1” to “4”. The questionnaire we ended up using is presented as following:

“Based on a literature review of driver behavior models, we have identified a list of features considered important for understanding this behavior. Please note, according to your judgment and experience, the involvement degree of each feature in accidents and bad driving habits.”
1) To what extent do the following features appear in traffic accident statistics? Responses are on a 4-point Likert scale, 1 = “rarely”, 2 = “every once in a while”, 3 = “sometimes” and 4 = “nearly all the time”.

<table>
<thead>
<tr>
<th>Feature</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad perception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtaking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive braking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not respecting signs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of vehicle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane changing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slips</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tailgating</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2) To what extent do the following features contribute to dangerous driving and affect driving performance? Responses are on a 4-point Likert scale 1 = “very mild”, 2 = “mild”, 3 = “moderate”, 4 = “severe”.

<table>
<thead>
<tr>
<th>Feature</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiousness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent-minded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensation seeking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forgetfulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5. Results of Driving-related features**

<table>
<thead>
<tr>
<th>Features</th>
<th>Data set 1: Literature questionnaires</th>
<th>Data set 2: CNPAC questionnaire</th>
<th>SAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class A</td>
<td>Class B</td>
<td>Class C</td>
</tr>
<tr>
<td>Anxiousness</td>
<td>3/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Carefulness</td>
<td>3/5</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Anger</td>
<td>2/5</td>
<td>0</td>
<td>1/5</td>
</tr>
<tr>
<td>Bad perception</td>
<td>2/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inattention</td>
<td>2/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overtaking</td>
<td>2/5</td>
<td>0</td>
<td>1/5</td>
</tr>
<tr>
<td>Speed</td>
<td>2/5</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Absent-minded</td>
<td>1/5</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Aggressive braking</td>
<td>1/5</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Aggressive expression</td>
<td>1/5</td>
<td>2/5</td>
<td>0</td>
</tr>
<tr>
<td>Not respecting signs</td>
<td>1/5</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>Patience</td>
<td>1/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Route planning</td>
<td>1/5</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>Sensation seeking</td>
<td>1/5</td>
<td>0</td>
<td>1/5</td>
</tr>
<tr>
<td>Turning</td>
<td>1/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Use of vehicle</td>
<td>1/5</td>
<td>2/5</td>
<td>0</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>0</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Distraction</td>
<td>0</td>
<td>1/5</td>
<td>0</td>
</tr>
<tr>
<td>Forgetfulness</td>
<td>0</td>
<td>2/5</td>
<td>0</td>
</tr>
<tr>
<td>Lane changing</td>
<td>0</td>
<td>0</td>
<td>1/5</td>
</tr>
<tr>
<td>Revenge</td>
<td>0</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>Slips</td>
<td>0</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>Tailgating</td>
<td>0</td>
<td>0</td>
<td>2/5</td>
</tr>
</tbody>
</table>
The main objective of this comparison is to observe the differences between driving behavior questionnaire features presented in literature and the parameters taken into account by the National Committee for Circulation Accidents Prevention (CNPAC) in Morocco. We calculate the sum of absolute differences to measure the statistical difference between the two data sets.

Results of the data analysis are presented in table 5. The P-value of Fisher’s exact test expresses the correlation between the findings we concluded from literature and the answers of road safety experts. This test is useful for small simple sizes, which applies in our case. The sum of absolute differences is a second test to uphold the values of the Fisher’s exact test. We are not going to consider a threshold for the p-value, but we rather identify the features with the highest p-value.

Not respecting signs, Turning, Alcohol consumption and Lane changing are the features with the highest values of SAD, which means there is a major divergence between data set 1 and data set 2. They are more ranked by the road safety professionals than by driver behavior questionnaires presented in literature. This difference can be explained by the nature of measurement studies of road users conducted by the CNPAC, they primarily depend on an information system that allow following basic law application and geographical monitoring of behavioral indicators, such as the distribution of speed according to different infrastructures, seat belt wearing for different types of road users, crossing light, respect of stop signs panels, helmet wearing etc. These features are selected in order to allow an understanding of accidents trends in different groups.

The results for Patience, Revenge and Distraction also show some significant difference, they have higher scores in data set 2, which is the CNPAC one. This difference demonstrates that these features have severe impacts on the number of accidents, even though they have low appearance in questionnaires.

V. CONCLUSION

This paper analyzes driver behavior from a qualitative approach; we collected the set of driving related features from behavioral questionnaires presented in literature. These features are collected from five questionnaires that have demonstrated their efficiency through many studies they were employed in. For each questionnaire we extracted the features considered in its items and ordered them by rate of appearance according to the Pareto principle. The outcome of this phase shows twenty three main features that have been considered as major aspects in the development of self-reported measurement for driver behavior, which are: Anxiousness, Carefulness, Anger, Bad perception, Inattention, Overtaking, Speed, Absent-minded, Aggressive braking, Aggressive expression, Not respecting signs, Patience, Route planning, Sensation seeking, Turning, Use of vehicle, Alcohol consumption, Distraction, Forgetfulness, Lane changing, Revenge, Slips, and Tailgating. These features will help create a generic model for driver behavior that includes every aspect of driving situations.

After assembling these features, we collaborated with the National Committee for Circulation Accident Prevention (CNPAC) of the Ministry of Transportation of Morocco. We developed our own questionnaire that includes all previous features, which experts of road safety rated according to their knowledge and experience. The questionnaire was targeted to collect data about the importance of each feature in the unfolding of projects conducted by the CNPAC each year to monitor the evolution of traffic accidents.

A limitation of the present study is the narrowed sample of our targeted population in the comparative study, with only eight professionals from the CNPAC. The results would have been more accurate if the targeted sample was diverse in terms of geographical distribution, and the questionnaire was validated by organizations from other countries than Morocco. Another limitation is the number of questionnaires we employed, if a second study is conducted with more questionnaires, results will definitely be different as the Pareto & ABC analysis is directly related to the content of the questionnaires.

From the results, we observed relevant differences between the two data sets. One of the most features to have significant difference is Distraction; it scored more in the CNPAC dataset than in the literature review. This finding shows that field researches give more attention to driver’s state of attention, which translates into distraction. Not respecting signs and Alcohol consumption have also higher rates in the CNPAC data set, which can be justified by their impact on accidents statistics.

This paper may be used as a reference for building a generic model for driver behavior, as we presented a set of important features that were both considered in self-report measurements researches and in field studies conducted by governmental organizations in Morocco.

REFERENCES


Authors’ Profiles

Kawtar Zinebi obtained an engineering degree in industrial management and production systems from MINES-RABAT school in 2015. She is currently pursuing a PhD in driver behavior modeling at Mohammadia School of Engineering.

Pr. Nissrine Souissi is a fulltime professor at the MINES-RABAT School, Morocco. She obtained a PhD in computer science from the University of Paris XII in 2006, France and an engineer degree in computer engineering from Mohammadia School of Engineers in 2001, Morocco. Her research interests include process engineering, business process management, databases, data lifecycle, smart data, hospital information system, and information system.

Pr. Kawtar Tikito is a fulltime professor at the MINES-RABAT School, Morocco. She obtained a PhD in computer science from the University Mohammed V of Rabat, Morocco and an engineer degree in computer engineering from Higher National School of Computer Science and Systems Analysis –ENSIAS, Morocco. Her research interests include system analysis, system optimization, information system and supply chain management.