ASSESSMENT OF DRIVING MENTAL MODELS AS A PREDICTOR OF CRASHES AND MOVING VIOLATIONS

A Thesis

by

GONZALO JAVIER MUÑOZ GÁLVEZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2011

Major Subject: Psychology

Assessment of Driving Mental Models as a Predictor of Crashes and Moving Violations

ASSESSMENT OF DRIVING MENTAL MODELS AS A PREDICTOR OF CRASHES AND MOVING VIOLATIONS

A Thesis

by

GONZALO JAVIER MUÑOZ GÁLVEZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Winfred Arthur, Jr.
Committee Members,	Stephanie C. Payne
	David J. Martin
Head of Department,	Ludy T. Benjamin, Jr.

May 2011

Major Subject: Psychology

ABSTRACT

Assessment of Driving Mental Models as a Predictor of Crashes and Moving Violations.

(May 2011)

Gonzalo Javier Muñoz Gálvez, B. S., Pontificia Universidad Católica de Chile

Chair of Advisory Committee: Dr. Winfred Arthur, Jr.

The purpose of the current study was to assess the efficacy of mental models as a predictor of driving outcomes. In contrast to more traditional measures of knowledge, mental models capture the configural property of knowledge, that is, an individual's understanding of the interrelationships that exist among critical concepts within a particular knowledge domain. Given that research has consistently shown the usefulness of mental models for the prediction of performance in a number of settings, it was hypothesized that the development of accurate driving mental models would also play an important role in the prediction of driving outcomes, especially in comparison to traditional measures of driving knowledge—such as the multiple-choice type tests typically required to obtain a driver license.

Mental models of 130 college students (52% females) between 17 and 21 years-old (M = 18.68, SD = 0.80) were analyzed and compared to a subject matter expert (SME) referent structure using Pathfinder. A statistically significant correlation was found for mental model accuracy and moving violations ($r = -.18$, $p < .05$), but not for at-

fault crashes. Evidence of incremental validity of mental models over commonly used predictors of moving violations (but not for at-fault crashes) was also found. Exploratory analyses revealed that driving knowledge, general mental ability (GMA), and emotional stability were the best predictors of mental model accuracy.

Issues related to the measurement of mental models were extensively addressed. First, statistically significant correlations between GMA and several mental model properties (i.e., accuracy scores, within participant similarity, and within participant correlation) suggest that challenges inherent to the task for eliciting mental models may influence mental model scores which, in turn, may lower mental model reliability estimates. Also, the selection of model components (i.e., terms) and the identification of the "best" reference structure for deriving mental model accuracy scores are undoubtedly critical aspects of mental model-related research. Along with illustrating the decisions made in the context of this particular study, some suggestions for conducting mental model-related research are provided.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

According to the National Highway Traffic Safety Administration (NHTSA) there were 5.8 million crashes in 2008 with an estimated 2.35 million people injured and 37,261 people killed which translates into a fatality rate of 1.27 per 100 million vehicle miles traveled (NHTSA, 2009a). In occupational settings, it has been estimated that transportation incidents account for 42% of workplace-related fatalities (Solis & Hall, 2009). Needless to say, the costs of traffic crashes are enormous, both for organizations and society.

Driving is a complex task that can be influenced by a myriad of individual differences, such as driving knowledge, information processing, demographic variables, exposure factors, and personality. An objective of the present study was to contribute to a further understanding of driving crashes and moving violations by introducing mental models to this stream of research. In contrast to more traditional measures of knowledge—which focus on the number of facts an individual can recognize or recall—mental models capture how knowledge is organized into meaningful conceptual structures to describe, explain, and predict future states of a given system (Rouse & Morris, 1986). Research has shown that the development of accurate mental models

---

This thesis follows the style of *Journal of Applied Psychology*.

plays a significant role in predicting performance on complex and dynamic tasks (Day, Arthur, & Gettman, 2001; Kraiger, Salas, & Cannon-Bowers, 1995), and thus, it is hypothesized that mental models should also play a significant role in the prediction of driving outcomes.

The validity of mental models as a predictor of crashes and moving violations must be assessed in conjunction with other predictors previously examined in the driving-related literature. Arthur and Day (2009) argued that a combination of individual differences for the prediction of driving crashes is consistent with the need to recognize the role of the whole individual in driving-related behaviors. They further demonstrated the usefulness of combining multiple predictor categories for predicting crashes and moving violations, over and above the usage of any single variable. In this sense, the specific objective of the current study was to investigate the incremental validity of mental models over and above other commonly used predictors in this domain, namely, demographic variables, exposure factors, information processing variables, and personality.

Although mental models have been successfully used as predictors of various performance criteria in academic and organizational contexts, this study represents, to the best of our knowledge, the first time mental models have been used as a predictor of driving outcomes. For this reason, a secondary objective of this study was to explore the relationship between mental models and the previously mentioned set of driving outcome predictors. This query aims to situate mental models among driving predictors' nomological network. Specifically, general mental ability (GMA), driving experience,

and personality were hypothesized to play an important role in the development of driving mental models. Hopefully, exploring these associations will shed light on the processes that underlie the development of mental models.

As is true with any other predictor variable, the validity of mental models as predictors of driving outcomes relies heavily on the quality of their measurement. As is well known, reliability represents the degree to which observed individual scores are indicative of their corresponding true scores (or latent dimension of interest). Specifically, reliability is the R^2-index associated with the regression of observed scores on true scores (Raykov & Marcoulides, 2011). Unless the association between an observed score and its correspondent true score is perfect ($R^2 = 1$), the observed predictor-criterion association will be less than the association between the latent factor and the specified criterion ($R^2 < 1$). Because of its relationship with error of prediction, predictor reliability must be taken into account when assessing the validity of any given predictor. Unfortunately, standard procedures for assessing reliability do not apply to mental model measurement without difficulty (at least as operationalized using Pathfinder). For instance, participants' mental models are not analyzable by means of internal consistency estimates based on correlational methods (e.g., Cronbach's alpha). Instead, mental model internal consistency must be assessed by means of *coherence scores*. Test-retest reliability can be assessed using mental model accuracy score correlations, within participant mental model correlations, and within participant similarity scores, but the relationship between these indexes is yet unclear.

Another measurement related concern is the decision process involved in selecting the best SME mental model to be used as a reference structure (i.e., a standard for comparing other mental models). Although a reference structure quality may be inferred by examining its properties (i.e., coherence scores and number of links), a rational examination of a mental model's quality may well contradict an empirical test of a reference structure's aptness for predicting a given criterion. For instance, a SME referent structure may show low internal consistency (i.e., low coherence score) but still be useful for predicting driving outcomes. Finally, a critical process in mental model measurement is that of sampling the right set and number of terms from the knowledge domain a mental model is purportedly representing. Because resultant mental models are fully determined by the components introduced by the researcher, selecting the terms may undoubtedly bias mental model assessment. The purpose of this study was not to give a definite answer for these or similar problems, and some of them may depend on following standard test development recommendations. However, the absence of clear-cut guidelines for carrying out some critical processes warrants some discussion. At any rate, the present study may serve to illustrate some of these issues (and how to handle them), and hopefully to engender further research on the topic of mental model measurement.

In the first portion of this thesis, a literature review of previously studied predictors of driving outcomes will be presented. Next, mental models will be defined—both theoretically and operationally—and introduced into this stream of research; the main hypothesis of this study will be introduced in this context.

2. EFFECTIVENESS OF CURRENT MODELS FOR PREDICTING CRASHES AND MOVING VIOLATIONS

The purpose of the following sections is to review some variables commonly used in the driving literature—including demographic variables, exposure factors, information processing variables, and personality—and examine their relationship with crashes and moving violations.

2.1 Demographic variables

Age and sex are probably the most commonly used predictors in the vehicle crash literature. First, there is evidence that most crashes occur among young male drivers. Twenty-three percent of all traffic fatalities occur among individuals 16 to 24 years old, the highest rate of involvement in fatal crashes on a per population basis (NHTSA, 2009b), and male drivers are involved in roughly three times more fatal crashes than women (NHTSA, 2009b). In spite of these figures, some researchers have argued that age and sex are proxies for other predictor variables more closely related to crash involvement, such as information processing and personality. Arthur et al. (1990) demonstrated that the correlation between age and crash rate was no longer significant after controlling for a test of selective attention, which suggests that a direct measure of information-processing ability is a better predictor than age per se. Williams and Carsten (1989) observed that the crash rate per miles driven for older drivers (65 years and older) is about the same as the crash rate per miles driven among younger drivers. The scarce attention to this problem among the elderly population is probably due to the fact that the

total number of crashes among older drivers is lower in comparison to the rest of the population, a result that is certainly misleading by itself unless one pays attention to the number of miles driven per week—another exposure factor that is reviewed in the next section. In general then, when making comparisons between populations it is important to keep in mind that the effect of the grouping variable on the criterion may be mediated by other predictor(s).

2.2 Exposure factors

Driving experience has been associated with driving performance improvement. Bellet, Bailly-Asuni, Mayenobe, and Banet (2009) demonstrated that experienced drivers outperform novice drivers in a situational awareness task. In this task, participants were requested to compare an image of a video sequence that was previously shown to them, and then indicate whether or not the image was modified (deletion or addition of a relevant element). Experienced drivers were able to detect 75% of the modifications (i.e., a pedestrian approaching the corner), whereas the novice drivers were able to identify only 59% of them.

With increased practice drivers gain driving-experience. However, this increase in the amount of driving is concomitant with more exposure, and subsequently a greater likelihood of being involved in traffic crashes and receiving more tickets. So, consonant with this, research typically shows that driving years is also positively associated with a higher number of crashes and moving violations. For instance, Arthur and Day (2009) found a positive correlation of .26 between driving years and at-fault crashes, and .43 with moving violations in a sample of college students.

Obviously, driving experience is also closely related to age, such that older drivers typically have more driving experience than younger drivers. Arthur and Day (2009) found a correlation of .84 between age and driving experience. At the same time, there is some evidence that as experienced drivers reach a certain age (65 years old), a reduction in cognitive capacities may hinder their driving performance in comparison with experienced drivers under 40 years old (Bellet et al., 2009). Nevertheless, even when differences in performance among different age groups have been found, they can often be attributed to within-group differences in information processing ability rather than age per se (Arthur, Barrett, & Doverspike, 1990). In other words, between group variance in performance associated with differences in age do not preclude the existence of within group differences on other variables relevant for predicting driving outcomes.

In general, the association between driving experience and driving outcomes must be examined with caution, especially in samples including elderly drivers in which driving experience may not be sufficient to compensate for the negative effect of reduced information processing capabilities relevant for competent driving. At the same time, it is reasonable to expect that as time elapses and driving experience accumulates, more practice will not necessarily produce better driving outcomes. In fact, once drivers have acquired the requisite knowledge and skill for driving, other predictors may play a more important role in the prediction of driving outcomes (e.g., personality). Finally, it is not unreasonable to posit that during the first stages of learning to drive, high ability individuals will acquire more knowledge and generate more accurate mental models than low ability individuals. Therefore, including driving experience as a predictor of

performance is relevant only to the extent that it is associated with an increase in driving knowledge which, in turn, may be mediated by other factors—specifically, GMA.

2.3 Information processing

Driving is a complex task that requires "perceiving, identifying, processing, and adequately responding to pertinent information (e.g., traffic lights, signs, pedestrians, other vehicles, etc.) in the environment" (Arthur & Day, 2009, p. 127). Because driving is complex, research should not underestimate the importance of having the necessary cognitive resources to successfully perform this task. Yet, it is not uncommon to find primary studies that fail to establish a consistent linkage between information-processing skills and crashes (see Arthur & Day, 2009).

GMA is one of the most successful information processing variables for predicting core (educational outcomes, occupational training, job performance) and peripheral criterion domains (such as crime, health risks, and poverty; Lubinski, 2000); however, its usefulness for predicting traffic crashes has not received empirical support in comparison to other information processing variables such as selective attention and perceptual style (Arthur, Barret, & Alexander, 1991). Arthur et al.'s (1991) meta-analysis identified selective attention and perceptual style as valid predictors of automobile crashes, with corrected mean *rs* of .26 and .15 respectively; in contrast, they found only marginally favorable results for cognitive ability, mean $r = .12$, 95% CI [.06, .18].

In their meta-analysis of dual-task driving studies, Horrey and Wickens (2006) demonstrated a significant deterioration of driving performance resulting from cellular

phone usage, an effect that was found to be equivalent for both hand-held and hands-free devices. Bellet et al.'s (2009) review also suggests that drivers in dual-task conditions show significant decreases in their ability to accurately represent their driving environment, therefore increasing their likelihood of being involved in a vehicle crash. These results run counter to a widespread belief in one's ability to perform multiple tasks while driving (e.g., texting and driving). For instance, it has been estimated that 11% of vehicles are being driven by someone using a hand-held or hands-free cell phone at any given daylight moment (NHTSA, 2009c).

In sum, research in this domain should not underestimate the complexity of driving-related tasks and the importance of having the necessary cognitive resources available to perform the driving task successfully. While the main purpose of this study is to test the efficacy of mental models—a measure of knowledge organization—as a predictor of driving outcomes, it is critical to keep in mind that information processing ability influences both skill acquisition and performance. As stated in the previous section, drivers with similar training may differ with regard to the time they need to develop sound and accurate driving mental models.

2.4 Personality

2.4.1 Five-factor model

The rationale for a multiple-predictor perspective to the prediction of driving outcomes is that crash involvement stems from two distinct sources of anomalous behavior, namely errors and violations (Arthur et al., 1991). While errors result from

failures in information processing ability, violations are more motivational, and reflect the preferences and choices the driver makes while driving. In turn, preferences and choices reflect a driver's personality.

Historically, a large number of personality variables have been used to predict crashes and moving violations. Fortunately, today there is a broad consensus on the use of the five-factor model (McCrae & Costa, 1987) for the study of personality and its relationship with relevant outcomes, especially in the field of personnel psychology and the prediction of job performance (Barrick & Mount, 1991; Schmidt & Hunter, 1998; Tett, Jackson, & Rothstein, 1991). Some of these personality variables, commonly used in organizational and personnel psychology research, have been successfully used in the prediction of crash involvement and moving violations. Clarke and Robertson's (2005) meta-analysis revealed that extraversion, conscientiousness, and agreeableness are valid predictors of crash involvement, with corrected mean validities of .24, .26 and .21, respectively.

Clarke and Robertson's (2005) results support the notion that high extraversion is associated with greater number of crashes on the road, but not with crashes in occupational contexts which, according to the authors, suggests that crash involvement in the case of extraverts may be related to monotonous or routine tasks. This evidence is consistent with the idea that extraverts have a somewhat lower level of vigilance, that is, the ability to sustain a high level of attention over long periods of time (Koelega, 1992; Schmidt, Beauducel, Brocke, & Strobel, 2004). Another argument that has been forwarded is that extroverts are high sensation seekers, and therefore have a tendency to

take greater risks when driving. Jonah's (1997) review reported a wealth of evidence supporting an association between sensation seeking and risky driving, including evidence suggesting that neurochemical differences in the brain as well as genetic factors underlie individual differences associated with risky driving.

In comparison to low conscientious individuals, those high in conscientiousness describe themselves as more careful, reliable, self-disciplined, persevering, and perceptive, among other things (McCrae & Costa, 1987). Not surprisingly, a number of studies have established a relatively consistent relationship between conscientiousness and driving outcomes (Arthur & Doverspike, 2001; Arthur & Graziano, 1996). For instance, Arthur and Graziano (1996) demonstrated a significant inverse relationship between conscientiousness and driving crash involvement in a sample of college students. Using a multivariate analysis of variance approach they found a small but significant mean difference between a group of participants who had one or more at-fault crashes in comparison to a group of individuals who reported no crashes in the same period of time. The same pattern of results was obtained for a sample drawn from a temporary employment agency. Bogg and Roberts' (2004) meta-analysis also confirmed that conscientiousness is associated with risky driving behaviors, such as drinking and driving, and speeding.

Low agreeableness individuals may be less able to cooperate with others and more prone to react aggressively to situations (Clarke & Robertson, 2005). The relationship between aggression—to a certain extent, the opposite of agreeableness—and crashes has been previously documented. An 8–year longitudinal study conducted in

New Zealand showed a significant relationship between some personality factors of the Multidimensional Personality Questionnaire (MPQ) and crashes among young adults (Gulliver & Begg, 2007). After controlling for driving exposure, they found that high levels of aggression (i.e., will frighten and cause discomfort to others), high levels of alienation (i.e., feels mistreated), and low levels of traditionalism (i.e., does not endorse high moral standards) were predictive of driving crashes. Consistent with these findings, West, Elander, and French (1993) found a positive association between number of crashes and mild social deviance, the tendency to engage in antisocial behavior from which harm to others would be a likely consequence (i.e., drive down the hard shoulder of a motorway when other lanes are jammed). Arthur et al.'s (1991) meta-analysis provided support for the linkage between number of crashes, general activity level (an aggression-related measure), and regard for authority, but only for a sub-population of professional drivers, not for the general population.

The efficacy of personality variables for predicting driving crashes has been established not only individually, but also when used in combination or as a set (Arthur & Day, 2009). According to Arthur and Day, the effect of a combination of personality variables (agreeableness, conscientiousness, emotional stability, extraversion, and openness) accounts for a statistically significant amount of variance over and above the variance explained by demographic and information processing variables when crashes or moving violations are used as criteria. Again, these results point to the importance and meaningfulness of using the "whole" person approach for predicting crashes and moving violations.

2.4.2 General and specific locus of control

Locus of control is a personality trait that represents the extent to which individuals perceive rewards or reinforcement as contingent upon their own behavior, skills, or internal dispositions (Rotter, 1966). If a person consistently interprets rewarding events as resulting from his/her own actions or internal dispositions, that person is said to have an internal locus of control. Conversely, if similar events are consistently perceived as the result of luck, fate, or some kind of external force, the person is said to have an external locus of control. The belief in internal or external locus of control reflects different kinds of learning paradigms which may play an important role during skill acquisition as well as performance. For instance, if a driver believes that crashes are the result of luck, he/she will regard safety-related driving behaviors as less important, and therefore, these behaviors will be less likely to be learned or enacted.

Arthur et al.'s (1991) meta-analysis obtained a mean *r* of .20 for locus of control and crash involvement. However, most of the studies included in this meta-analysis used Rotter's (1966) locus of control scale, a measure that reflects a very broad or general disposition to attribute rewards to internal or external causes. The lack of specificity in Rotter's operationalization of locus of control has been discussed in the context of the frame-of-reference literature, which emphasizes the importance of imposing a specific frame of reference on test takers to improve the validity of personality scales (Lievens, De Corte, & Schollaert, 2008). A recent meta-analysis by Qiang, Bowling, and Eschleman (2010) revealed that a domain-specific measure of locus of control (work locus of control) was correlated with work-related criteria—such as job attitudes, job

performance, and life satisfaction, among other things—over and beyond the effect of general locus of control.

Whereas the frame-of-reference approach has been successfully used to improve the validity of personality tests in the context of work-related behaviors, its value in the context of driving-related behaviors has received mixed support. Using a postdictive design, Montag and Comrey (1987) obtained a multiple correlation of .38 between fatal crashes and the Montag Driving Internality and Driving Externality (MDIE) scale, a measure of locus of control tailored specifically to driving behavior. In contrast to the initial findings of Montag and Comrey, Arthur and Doverspike's (1992) study showed that although a statistically significant correlation between MDIE and not-at-fault crashes was found, this correlation was in the opposite direction to that typically hypothesized by the locus of control-crash involvement research ($r = -.16$, $p < .05$).

In sum, whereas personality testing research tends to support the frame-of-reference approach—in particular, a domain-specific measure of locus of control in the context of work-related behaviors—the advantage of using modified scales to measure locus of control within the context of driving-related behaviors has yet to be determined.

2.5 Driving knowledge

State regulations typically require driver license applicants to take a knowledge test, or provide evidence that the person has completed a valid driver education course. Moreover, under certain circumstances, drivers may dismiss traffic tickets from driving records by taking a defensive driving course. However, empirical assessments of the relationship between knowledge and traffic crashes suggest that declarative knowledge

is not a valid predictor of traffic crashes (Arthur & Doverspike, 2001). This empirical finding is consistent with Ackerman's (1988) learning acquisition model, in which declarative knowledge predicts skill acquisition in early stages, but not in later stages where other individual differences (such as perceptual speed ability and psychomotor ability) play a larger role. Ackerman (1988) argued that "broad intellectual abilities appear to correlate substantially with initial task performance, but these correlations diminish as skills are acquired. Furthermore, some perceptual-motor abilities that initially show small correlations with performance increase during practice" (p. 289). In other words, there are both theoretical reasons and empirical evidence suggesting that declarative knowledge is not a valid predictor of traffic crashes.

3. DRIVING MENTAL MODELS AS A PREDICTOR OF CRASHES AND MOVING VIOLATIONS

The complex and dynamic nature of knowledge requires the use of an inclusive taxonomy that captures the different facets of knowledge, and identifies the appropriate assessment techniques to represent and evaluate distinct learning outcomes. Kraiger, Ford, and Salas (1993) described three general categories of learning outcomes—cognitive, skill-based, and affective. These authors identified knowledge organization as one of the three learning categories comprising the cognitive domain (along with verbal knowledge and cognitive strategies). Knowledge organization is a higher order level of knowledge that captures the relationships that exist between a set of concepts within a given knowledge domain. In turn, a mental model is a structural representation of an individual's knowledge organization that allows us to capture this configural property of knowledge. The present study represents an attempt to expand the cognitive-predictor domain in the driving-related literature by introducing mental models, a cognitively-based measure of learning for assessing how knowledge is structured. Consonant with Kraiger et al., it is argued that as expertise in the driving domain develops, traditional knowledge tests would not be as sensitive as mental models in discriminating between individuals with different levels of proficiency, and therefore, mental models will increase the ability to predict driving outcomes. In the next section, the theoretical underpinnings of mental model research as well as the operationalization of this

construct will be discussed along with the rationale for using mental models as a potential predictor of crashes and moving violations.

3.1 Mental models: Conceptual delimitation and research findings

A mental model is a network of associations between concepts in an individual's mind. Mental models have been described as a form of *intuitive knowledge* that serves as a frame of reference for interpreting the world which forms the bases for reasoning and working with problems (Johnson-Laird, 1983; Kraiger et al., 1993; Rouse & Morris, 1986). Accordingly, mental model assessment techniques elicit the configural property of knowledge—in contrast to traditional knowledge tests which measure the amount of information an individual can recognize or recall.

Mental models can be measured using multiple techniques, although structural assessment is the most common approach for assessing knowledge structures (Kraiger et al., 1993). Structural assessment implementation involves three steps: (a) knowledge elicitation, (b) knowledge representation, and (c) evaluation of an individual knowledge representation (Goldsmith, Johnson, & Acton, 1991). Knowledge elicitation refers to the technique used to draw relatedness judgments between pairs of concepts, such us similarity ratings, card sorting, or concept mapping. Once the relationship between every model component has been determined, this information can be arranged in the form of a proximity matrix, in which each cell value corresponds to the relatedness between a single pair of concepts. Knowledge representation refers to the technique or scaling procedure used to derive an individual's cognitive structure. A computer program can be used to draw a spatial representation of a mental model, in which each node represents a

component of the model (e.g., a concept), and a link between nodes has a weighted value determined by the distance between the model components.

Resultant mental models can then be evaluated in terms of similarity and accuracy scores. Similarity is typically used in the context of team research, and refers to the extent to which team members share an understanding of the team's environment (see Mohammed, Ferzandi, & Hamilton, 2010). In contrast, accuracy scores represent the extent to which an individual's mental model approximates an expert model, a referent structure that best reflects the "true" structure of the domain (Acton, Johnson, & Goldsmith, 1994). Regardless of which mental models are being compared (i.e., team members' mental models or novice versus expert mental models), the metric used to represent mental model similarity and accuracy is the same. Previous research has established that mental model comparisons using Pathfinder are better than other scaling algorithm methods, such as proximity data and multidimensional scaling (Goldsmith & Johnson, 1990; Goldsmith et al., 1991). Specifically, Goldsmith and Johnson (1990) advocate the use of a metric known as closeness (*C*) as the best way to operationalize similarity between two mental models (e.g., between a novice and an expert mental model). *C* is roughly equal to the ratio of the number of common links between two networks divided by the total number of links in both; it varies from 0 to 1 with one representing a perfect match between two mental models (Day, Arthur, & Gettman, 2001; Kraiger, Salas, & Cannon-Bowers, 1995).

Given that the purpose of this study was to assess the efficacy of mental models as predictors of driving outcomes—an individual task—only accuracy scores will be

discussed henceforth. Again, accuracy scores reflect a participant's similarity not to other participants, but to a referent structure which is "some ideal organization that best reflects the structure of the domain" (Acton, Johnson, & Goldsmith, 1994, p.303). Since cognitive structures should become more like this ideal structure as experience in the domain increases, a customary practice is to ask an expert (or panel of experts) to generate a mental model of the domain, which in turn is used as the point of comparison for novices' mental models. Therefore, accuracy represents the degree of similarity with a reference structure, and indicates how closely a specified mental model resembles a "true" state of the world. Notwithstanding, it is important to recognize that it is possible to find SMEs who have "equally good yet different mental models for any given content domain" (Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005, p. 39). This phenomena, known as equifinality, occurs when alternative mental models are equally effective in predicting task performance. For instance, in making a left turn, person A may choose to make a full stop at the intersection and wait until the red light changes to green, and then beats the oncoming traffic through speed. Under similar circumstances, person B may drift slowly after entering the crossroads and make a left turn without making intermediate stops (see Bellet et al., 2009; *opportunistic* vs *dynamic regulation* strategies). Both drivers are equally successful in making left turns, but they have different mental models for performing this task. The idea of "braking" and "left turn" should be closely related in person A's mental model (i.e., stop before turning left), whereas in case of driver B, "speed control" should be more closely related to "left turn" (i.e., slow down as entering a crossroad). Also, it should be noted that the use of the term

equifinality does not preclude the use of the term accuracy; it just means that the task permits multiple paths to effective performance.

Research has shown the usefulness of comparing novice versus expert referent structures for the prediction of performance in academic (Davis, Curtis, & Tschetter, 2003; Goldsmith et al., 1991) and organizational contexts (Davis & Yi, 2004). Simply put, as accuracy scores increase individuals tend to perform better. Likewise, several studies have shown that *C* scores change as a function of training. For instance, Goldsmith and Johnson (1990) showed that structural assessments of mental models can be used to determine how a student's configural representation of a subject matter becomes increasingly similar to an expert model over the course of instruction. Also, Kraiger et al. (1995) showed that *C* scores were positively correlated with performance in a tactical decision making task ($r = .63$), but only for a subgroup who received advance organizers (information on task goals and training objectives) prior to training. Day et al. (2001) demonstrated that the similarity between trainee's knowledge structure and an expert model was correlated with skill acquisition, skill retention, and transfer. Furthermore, they demonstrated that accuracy scores mediated the association between GMA, and skill retention and transfer. Finally, Smith-Jentsch, Campbell, Milanovich, and Reynolds (2001) showed that mental models were more accurate among higher ranking navy personnel, and showed that teamwork mental models were amenable to change through a computer-based training intervention (see also Edwards, Day, Arthur, & Bell, 2006).

3.2 The nature of driving mental models and its relationship with driving outcomes

Driving mental models can be defined as representational schemas for driving-related tasks. Bellet et al. (2009) describe car driving as a dynamic process that requires constant adaptation from the driver. According to this view, driver mental models are formed under specific circumstances, determined by both the driving situation and the needs and intentions of the driver. That is, mental models emerge from an iterative perception-action cycle in which an active observer is constantly building models of the task in accordance with his/her goals. Once these mental representations are formed, they can be stored in long term memory and reactivated in future performance of similar tasks. Bellet et al. introduced the term immergence "to describe the sedimentation phenomenon of explicit thought in implicit knowledge" (p. 1209). This concept is similar to Anderson's (1993) notion of compilation, in that once a mental representation is formed (or emerges) and put into practice repeatedly over time, it becomes an automated process that is no longer under the person's explicit awareness or control—or, the mental representation immerges in deep-cognition.

A reason for using mental models to assess driving knowledge is the notion that expertise in this domain may require not only the acquisition of knowledge, but an understanding of the interrelationships between critical driving components, a level of comprehension that runs parallel to skill development (Anderson, 1982). An assumption of this model is that the linear association between declarative knowledge and driving outcomes is not monotonic, such that after a cut-off score has been surpassed the correlation between declarative knowledge and driving outcomes would no longer

remain statistically significant. In other words, as expertise in this domain develops, traditional knowledge tests would not be as sensitive as mental models in discriminating between individuals with different levels of proficiency (Kraiger et al., 1993). In support of this reasoning, Kraiger et al. (1995) did not find significant differences between groups under different experimental conditions when comparisons were made on the basis of a multiple-choice declarative knowledge test (because of this measure's range restriction effect). Conversely, they found that experimental and control training groups differed significantly in terms of their mental models; a finding which indicates that mental models differentiate better at the high end of the performance range.

Consonant with the extant body of literature, the following hypothesis were tested:

Hypothesis 1a: Mental model accuracy will predict at-fault crashes.

Hypothesis 1b: Mental model accuracy will predict moving violations.

Hypothesis 2a: Mental model accuracy will predict at-fault crashes better than a traditional measure of driving knowledge (i.e., a driving knowledge test)

Hypothesis 2b: Mental model accuracy will predict moving violations better than a traditional measure of driving knowledge (i.e., a driving knowledge test)

Hypothesis 3a: Mental model accuracy will account for variance in crashes over and above other commonly used predictors of driving outcomes.

Hypothesis 3b: Mental model accuracy will account for variance in moving violations over and above other commonly used predictors of driving outcomes.

As a second line of inquiry, the relationship between driving mental models and other commonly used predictors of driving outcomes was also explored. First, if mental models develop as experience accrues, then individuals with more driving experience (i.e., more years driving, more miles driven per week) should develop more accurate mental models. Also, as pointed out by Bellet et al. (2009), mental models carry information about the driver's value judgments of his/her own driving, which in turn are a reflection of his/her personality.[1] If this later claim is true—that mental models are also shaped by driver's personality—then it follows that people with different personalities should also develop different driving mental models. This later claim is consistent with Ackerman's notion of trait complexes, which are non-ability traits that presumably facilitate or impede the acquisition of domain knowledge (Ackerman, 2003, 2007; Ackerman & Heggestad, 1997). Finally, there is evidence that GMA facilitates skill acquisition. In this sense, driving mental model accuracy (i.e., a learning outcome) should be predicted by GMA. In a previous study, Day et al. (2001) showed that the correlation between mental model accuracy and GMA was statistically significant, with

[1] Bellet et al. (2009) refer to this process as *reflexivity*, which comprises the meta-cognitions about the behavior performed during driving, and a value judgment dimension about one's acts that accompanies the development of mental representations. These meta-cognitions about driving-related behaviors are carried into deep-cognition during the immergence process.

a mechanically combined referent structure yielding better results ($r = .45$, $p < .001$) than a consensus referent structure ($r = .29$, $p < .01$).

In accordance, the following hypotheses were tested:

Hypothesis 4: Driving experience will be positively associated with mental model accuracy.

Hypothesis 5: Personality variables typically associated with driving outcomes (i.e., conscientiousness, extroversion, and agreeableness) will be associated with mental model accuracy.

Hypothesis 6: GMA will be positively correlated with mental model accuracy.

4. METHOD

4.1 Participants

The sample consisted of 130 individuals (52% female) between 17 and 21 years old (*M* = 18.68, *SD* = 0.80). They were undergraduate students from a southwestern university and were recruited as part of a larger study on teams. All measures for the present study were collected at different points during a proctored lab session that lasted five hours.

4.2 Measures

4.2.1 Raven's Advanced Progressive Matrices (APM)

GMA was operationalized as scores on the short form of the APM (Arthur & Day, 1994; Arthur, Tubre, Paul, & Sanchez-Ku, 1999) which consists of 2 practice items and 12 test items. This measure is regarded as a test with a low level of culture loading, and it is considered the purest available measure of fluid intelligence (Raven, 1989, 2000). Arthur et al. reported a 1−week test-retest reliability of .76 for this measure.

4.2.2 Driving behavior questionnaire

Total crashes and number of tickets were collected via a self-report measure (Arthur & Doverspike, 1992). Specifically, participants were asked to list all crashes in which they were at-fault and not-at-fault each year since 2005 and up to 2009 ("How many At-fault accidents were you in?", "How many not at-fault accidents?"). They were also asked to report the number of tickets or moving violations they had received by year

for the same period of time ("How many tickets or moving violations did you receive?"). The same questionnaire was used to collect information about driving experience (number of years driving rounded to the nearest year) and number of miles driven per week. On average, participants had been driving for 3.02 years (*SD* = 1.14, min = 1, max = 6). Table 1 displays participants' type of driving license and type of vehicles they usually drive. Participants also responded to a question about speed limit abiding. Specifically, they were asked "On average, how many miles per hour under or over the speed limit you drive?".

4.2.3 Driving knowledge test

Driving knowledge was assessed via an 18–item four-alternative multiple-choice test. Items for this test were developed from the Texas Drivers Handbook and reflect knowledge required in order to obtain a Texas driver's license. The average number of correct responses for this measure was 12.49 (*SD* = 2.00, min = 6, max = 17).

4.2.4 Five-factor model

Goldberg's (2006) 50–item measure was used to operationalize the five-factor model personality variables—extroversion, agreeableness, conscientiousness, emotional stability, and openness. Each factor was assessed with 10 Likert-type items, with participants rating the extent to which each statement was descriptive of them (1 = *very inaccurate*, 5 = *very accurate*). Internal consistency estimates for the factors ranged from .80 to .89.

Table 1
Descriptive Statistics for Driving Questionnaire

	N	*%*
Type of License		
No response[a]	2	1.54
Class A	7	5.38
Class B	6	4.62
Class C	111	85.38
Class M	4	3.08
Total	130	100.00
Type of Vehicle		
No response	1	0.77
Passenger car	49	37.69
Pick-up truck	26	20.00
Sport-utility vehicle (SUV)	40	30.77
Motorcycle	5	3.85
Van	8	6.15
Commercial vehicle (e.g., 18–wheeler, bus)	1	0.77
Total	130	100.00

Note. Class A = for operating vehicles which tow trailers or other vehicles over 10,000 pounds; Class B = for operating vehicles over 26,001 lbs; Class C = for operating vehicles under 26,001 lbs. that would normally not require a commercial driver license; Class M = for operating motorcycles.
[a] Although these participants did not provide their license type, based on their responses to other questions of the driving questionnaire, it was safe to assume that both of them were actual drivers.

4.2.5 General and specific locus of control

Rotter's (1966) internal-external scale was used to measure general locus of control, and was scored such that higher scores reflect an external locus of control and lower scores an internal locus. The internal-external scale consisted of 23 pairs of statements,[2] and participants were instructed to select one statement from each pair that better reflects their personal beliefs. Previous studies have reported moderate to high internal consistency estimates, ranging from .69 to .76 (Rotter, 1966). Coefficient alpha for our sample was .67 for general locus of control.

Montag and Comrey's (1987) driving internality and driving externality measure was used to operationalize specific locus of control. Each scale (internality and externality) consisted of 15 statements and participants were instructed to indicate the extent to which they agreed with each statement (1 = *disagree very much*, 6 = *agree very much*). Although the original measure takes a multi-dimensional approach, we combined both scales to yield a single score of specific locus of control, such that higher scores reflect an external locus of control and lower scores an internal locus. The coefficient alpha for the combined 30-item measure was .77.

4.2.6 Mental models

Pathfinder (Schvaneveldt, 1990), a structural-assessment-technique-based program, was used for the elicitation, analysis, and comparison of driving mental models. The 12 concepts (or terms) used as the model components were extracted from a

[2] The original scale consists of 23 pairs of statement, but because of a technical problem we were able to use only the first 18 statements.

driving education text book. Participants were asked to make judgments about the relatedness of all possible pairs of concepts. They were instructed to think about the concepts as they relate to driving safely. Each concept was presented at the center of a bull's eye type diagram and participants were instructed to drag-and-drop the remaining 11 concepts around a target concept (Figure 1). Relatedness ratings were based on the distance between a given concept and the target (0 = *less related or unrelated*, 4 = *synonyms*). The same procedure was repeated until every concept was positioned as the target concept. After relatedness ratings were collected, Pathfinder was used to render network structures. Networks were derived using two parameters, r and q, to determine how network distance is calculated (r was set to infinity and q was set to equal the number of concepts minus one).

A measure of internal consistency, known as coherence score, was computed for each participant dataset. Coherence scores range from 0 to 1, and are based on the assumption that relatedness between a pair of items can be predicted by the relations of those items to other items in the set. Low coherence values (< .20) may indicate that raters have little or no expertise in the domain, or it may indicate that they did not take the task seriously. In order to estimate this measure's reliability, participants' mental models were assessed on two different occasions with a retest interval of approximately 40 minutes.

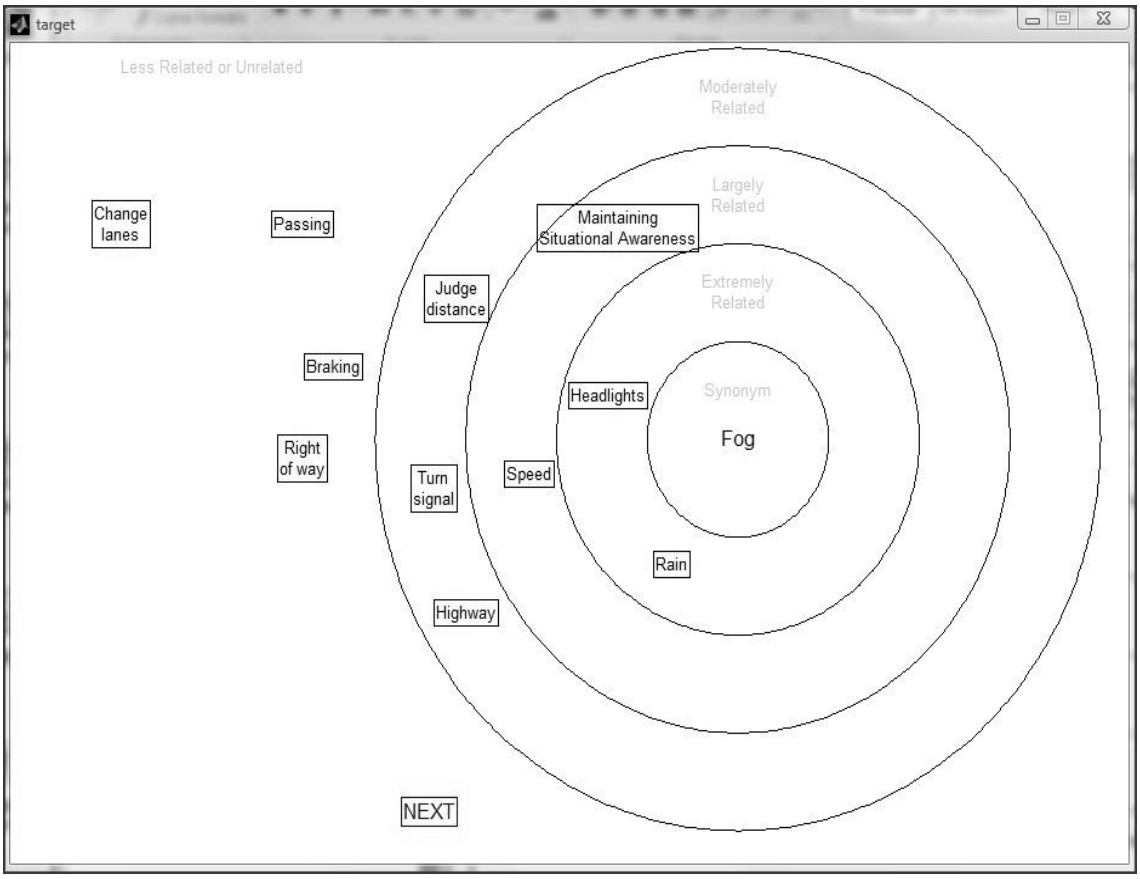

Figure 1. Screen capture of a participant's view of the bull's eye diagram for arranging mental model terms. The focal term ("Fog") is at the center of the bull's eye while the remaining terms are being dragged and dropped around the focal term to indicate their degree of relatedness (synonyms, extremely related, largely related, moderately related, and less related or unrelated [outside the bull's eye])

5. RESULTS

5.1 Descriptive statistics

Table 2 displays the number of crashes and moving violations reported by study participants. Remarkably, 35 participants (26.92%) reported being involved in at least one crash in which they were at-fault during the last five years, whereas 28 (21.53%) were involved in a car crash in which they were not-at-fault. In comparison to at-fault and not-at-fault figures, the percentage of individuals who received at least one ticket during the same period of time was somewhat higher (31.53%).

Table 2
Crashes and Moving Violations Reported by Study Participants

		N	%
Total At-fault Crashes			
	0	95	73.08
	1	21	16.15
	2	13	10.00
	4	1	0.77
Total Not-at-fault Crashes			
	0	102	78.46
	1	20	15.38
	2	6	4.62
	3	2	1.54
Total Moving Violations			
	0	89	68.46
	1	24	18.46
	2	11	8.46
	3	4	3.08
	4	1	0.77
	6	1	0.77

Note. $N = 130$. Data from 2005 to 2009.

As shown in Table 3, the correlation between sex and at-fault crashes was statistically significant, with females having more crashes than males, $r = -.18$, $p < .05$. The correlation between age and number of tickets was also statistically significant; however, age was also positively correlated with driving years and miles driven per week, so further analyses were conducted to partial out the effect of both variables separately. After controlling for driving years and miles driven per week the association between age and moving violations was no longer statistically significant. Conscientiousness was associated with at-fault and not-at-fault crashes, as well as moving violations, *rs* between –.17 to –.24.

5.2 Mental models accuracy scores

Accuracy scores were obtained by comparing each participant's mental model to a mechanically combined SME mental model. Five police officers from a mid-size municipal police department were recruited to serve as SMEs. All of them were currently active police officers with 4 to 25 years-of-service ($M = 18.40$), with a diverse driving training background in areas such as defensive driving, crash investigation, crash involvement, crash reconstruction, and traffic enforcement, among others. The same procedure described above for collecting participants' relatedness ratings was followed.

Table 3

Descriptive Statistics and Study Variable Intercorrelations

	Criteria			Demographic		Exposure			GMA	Knowledge		
Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1. At-fault crashes												
2. Not-at-fault crashes	.29***											
3. Tickets	.34***	.21*										
4. Sex[a]	–.18*	–.08	–.11									
5. Age	.07	.12	.21*	.23**								
6. Driving years	.15	.09	.22*	.17	.50***							
7. Miles per week	.17	.07	.35***	.07	.21*	.23*						
8. Speed	.15	.01	.36***	–.17	.08	.00	.17					
9. GMA	.07	.20*	–.04	.10	.03	.08	.05	.09				
10. Driving knowledge test	.04	.03	.05	–.02	–.07	.12	.28**	.09	.18*			
11. Coherence adm. 1	.06	.07	.04	–.05	.05	.04	–.03	.10	–.03	.04		
12. Accuracy adm. 1	–.01	.04	–.16	.09	.00	–.02	–.13	–.07	.31***	.07	.12	
13. Coherence adm. 2	.06	–.03	–.07	.11	–.01	–.04	–.01	–.01	.04	.16	.31***	–.01
14. Accuracy adm. 2	.03	–.05	–.18*	.01	–.09	–.15	–.03	–.03	.24**	.25**	–.08	.49***
15. Within participant sim	.16	.04	–.08	.03	–.02	–.03	.02	.07	.25**	.23**	.11	.51***
16. Within participant corr	.08	.08	–.10	–.12	–.12	–.16	–.14	.08	.24**	.11	.19*	.35***
17. Extroversion	.16	–.15	.09	–.05	–.01	.11	.02	.11	–.01	.05	.06	–.08
18. Agreeableness	.07	.02	–.05	–.21*	–.06	–.06	–.07	–.09	–.05	.07	.14	.00
19. Conscientiousness	–.21*	–.17*	–.24**	–.12	–.15	–.07	–.17	–.01	.16	–.08	.10	.23**
20. Emotional Stability	–.13	–.08	–.10	.25**	.04	.04	–.04	–.13	.16	.15	.02	.28**
21. Openness	–.04	.02	–.17	.12	.01	.02	–.12	–.12	.19*	.07	–.01	.09
22. General LoC	.03	.13	.05	–.17	–.07	.00	–.02	.10	.04	–.01	–.04	.05
23. Specific LoC	–.06	.03	.05	–.13	–.07	–.07	–.10	–.04	–.06	–.01	.10	–.04
M	0.39	0.29	0.52	0.49	18.69	3.02	78.57	5.19	8.22	12.49	.20	.17
SD	0.73	0.63	0.97	0.50	0.80	1.14	109.35	4.24	2.21	2.00	.25	.08

Note. $N = 130$. GMA = general mental ability; LoC = locus of control; adm. = administration; sim = similarity; corr = correlation.
[a]Dummy codes for sex are female = 0, and male = 1. *$p < .05$. **$p < .01$. ***$p < .001$. All tests are two-tailed.

Table 3 (continued)

	Knowledge					Personality					
Variable	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*
1. At-fault crashes											
2. Not-at-fault crashes											
3. Tickets											
4. Sex											
5. Age											
6. Driving years											
7. Miles per week											
8. Speed											
9. GMA											
10. Driving knowledge test											
11. Coherence adm. 1											
12. Accuracy adm. 1											
13. Coherence adm. 2											
14. Accuracy adm. 2	.09										
15. Within participant sim	.13	.57***									
16. Within participant corr	.22*	.38***	.65***								
17. Extroversion	.03	.02	.06	.02							
18. Agreeableness	.16	–.06	–.05	.08	.34***						
19. Conscientiousness	.03	.05	–.03	.04	–.06	.04					
20. Emotional Stability	.05	.21*	.16	.15	.07	.10	.08				
21. Openness	.02	.07	.15	.09	.29***	.22*	.03	.15			
22. General LoC	–.06	–.06	–.06	.05	–.10	–.07	–.08	–.26**	.02		
23. Specific LoC	.00	–.14	.00	.08	–.02	.07	–.06	–.20*	.05	.23**	
M	.19	.17	.24	.44	34.24	38.98	34.75	32.05	36.59	8.51	3.45
SD	.29	.09	.15	.22	7.18	5.39	5.81	6.93	5.30	3.18	0.43

Note. $N = 130$. GMA = general mental ability; LoC = locus of control; adm. = administration; sim = similarity; corr = correlation.
[a]Dummy codes for sex are female = 0, and male = 1. *$p < .05$. **$p < .01$. ***$p < .001$. All tests are two-tailed.

Except for SME 1, all SMEs' mental models had coherence scores over the .20 cutoff point (Table 4). Along with the low coherence score, SME 1 had the largest number of links between nodes (42) which can be an indicator of low ability to discriminate among concepts. Network similarity between SMEs was assessed using *C* scores.[3] Most *C* scores obtained were statistically significant ($p < .05$), which means that the number of links in common between a pair of networks is greater than what it would be expected by chance.

Table 4
Coherence Scores and Number of Links of Each SME Mental Model, and Two Measures of Between-SME Mental Model Similarity (Number of Common Links and C Scores)

	Coherence	SME 1	SME 2	SME 3	SME 4	SME 5
SME 1	-.02	(42)	31	22	29	22
SME 2	.26	.29**	(34)	20	25	16
SME 3	.28	.15**	.21**	(26)	20	12
SME 4	.53	.17**	.19**	.18**	(36)	16
SME 5	.27	.22**	.13**	.09	.10	(23)

Note. SME = Subject matter expert. Coherence scores lower than .20 indicate that mental model is not coherent. The number of links of each SME mental model are displayed in parenthesis. Number of common links and *C* scores are displayed above and below the diagonal respectively. *t*-tests are based on corrected similarity scores with $p < .05$ indicating that networks are similar.
*$p < .05$. **$p < .01$.

In order to identify the best referent structure, every combination of SME's mental models was analyzed and compared—an average mental model of all four SME's, as well as average network of every possible group of two and three SMEs' mental models. The mean coherence score of every combination taken together was .30

[3] In the Pathfinder program referred to as *corrected similarity scores*.

(*SD* = .16). The mean coherence scores for groups of two SME mental models (M_c = .41, *N* = 6) was higher than groups of three SME mental models (M_c = .16, *N* = 4), and somewhat higher than the combined network of all SMEs together (*C* = .34).

As it can be seen in Table 5, the correlation between mental model accuracy measured at Administration 1 and 2 and at-fault crashes did not yield any statistically significant relationships. For not-at-fault crashes there are several significant correlations but in the opposite direction, that is, high mental model accuracy was associated with relatively high not-at-fault crashes rate. The correlations between mental model accuracy and moving violations are generally low but they tend to increase as the number of SMEs increases, that is, when an additional SME mental model is added to the referent structure the correlation with criteria is improved (from .02 to –.14, excluding a combined network of all four SMEs wherein the correlation with criteria drops to –.04). At the same time, changing the number of SMEs for creating the mechanically combined referent structures also impacts coherence scores. At first, coherence scores improve, on average, from .27 (*SD* = .20) to .41 (*SD* =.07); but then, when a third network was added to the referent structure, the average coherence score drops to .16 (*SD* = .15).

Table 5

Different Individual and Mechanically Combined SMEs Mental Models Properties and Their Correlations to Criterion Variables

SME	Coherence	Number of links	Administration 1			Administration 2		
			At-fault crashes	Not-at-fault crashes	Moving violations	At-fault crashes	Not-at-fault crashes	Moving violations
SME 1	–.02	42	–.06	.19*	–.06	.03	.13	–.07
SME 2	.26	34	–.05	.18*	–.14	.13	.15	–.04
SME 3	.28	26	–.06	–.02	–.06	.08	.12	.17
SME 4	.53	36	.01	.22*	–.01	.14	.13	.08
SME 5	.27	23	.03	.03	–.13	–.02	–.02	–.14
Mean	.34	32.20	–.03	.12	–.08	.07	.10	.00
SD	.20	7.69	.04	.11	.06	.07	.07	.12
SME 23	.34	19	–.06	.06	–.08	.06	.10	.09
SME 24	.46	24	–.02	.21*	–.09	.07	.08	.02
SME 25	.40	19	–.01	.04	–.16	.03	–.05	–.18*
SME 34	.51	19	.02	.14	.06	–.09	–.04	–.16
SME 35	.33	15	–.02	.20*	–.03	–.09	.06	.00
SME 45	.41	15	.02	.14	–.08	–.07	.02	–.12
Mean	.41	18.50	–.01	.13	–.06	–.01	.03	–.06
SD	.07	3.33	.03	.07	.07	.08	.06	.11
SME 234	.16	26	.04	.08	–.02	.00	–.03	–.13
SME 235	.37	25	–.04	.18*	–.08	–.02	.11	–.12
SME 245	.01	26	.03	.19*	.04	–.01	.14	–.10
SME 345	.10	25	.06	.01	.09	–.09	–.09	–.22*
Mean	.16	25.50	.02	.11	.01	–.03	.03	–.14
SD	.15	0.58	.05	.09	.07	.04	.11	.05
SME 2345	.34	19	.00	.21*	–.01	.02	.09	–.04
Total mean	.30	24.56	–.01	.13	–.05	.01	.06	–.06
SD	.16	7.48	.04	.08	.07	.07	.08	.11

Note. A combined mental model of all five SME is not provided because the mental model of one of them (SME 1) did not meet the .20 cut-off score for coherence. *p <. 05, two tailed. N = 130.

Because the similarity between SMEs's mental models was not perfect (see Table 4) aggregating them to create a referent structure may introduce some inconsistencies that may negatively impact coherence scores. It is possible that this set of experts hold accurate mental models of driving safely but are also inconsistent between each other. If this is the case, then adding them up would create a distorted picture of the domain. At the same time, adding more expert knowledge to the model should provide a broader picture of the content domain which may translate into better predictor-criterion correlations. The fact that higher predictor-criterion estimates were obtained for referent structures based on combined SME mental models in comparison to estimates obtained from individually based reference structures seems to support this later claim. Low coherence scores may be just an undesired "side effect" of expanding the knowledge domain.

Only a combined network of SMEs 2 and 5 (henceforth SME25), and a combined network of SMEs 3, 4, and 5 (henceforth SME345) displayed statistically significant correlations with criteria in the expected direction. However, the referent structure SME25 will be used to test the main hypotheses of this study instead of SME345 based on the following reasoning. First, SME345 is not coherent (C = .10). Although coherence is not correlated with a referent structure's ability to predict crashes and moving violations (see Table 6), having an incoherent network as referent structure makes the results more difficult to interpret. For instance, highly accurate mental model may also be highly incoherent mental models.

Table 6
Correlations Between All Referent Structures Psychometric Properties and Driving Outcomes

	Coherence	*Number of links*
Administration 1		
At-fault crashes	–.06	–.28
Not-at-fault crashes	.14	.22
Moving violations	–.21	–.03
Administration 2		
At-fault crashes	.14	.58*
Not-at-fault crashes	–.09	.48
Moving violations	.24	.18

Note. *$p < .05$, two-tailed. $N = 16$. These results should be interpreted with caution because some combined SME mental models share raters.

Second, SME345 had a larger number of links which may indicate less differentiation between concepts. (At the same time, as shown in Table 6, a higher number of concepts was associated with higher crash rate at Administration 2.) Lastly, SME3 mental model was associated with moving violations in the opposite direction whereas, individually, both SME2 and SME5 mental models were associated with moving violations in the expected direction (although neither of these correlations were statistically significant).

Since the primary focus of this study was to predict driving-related criteria, an average network of SME2 and SME5 (Figure 2) was used as the referent structure to derive participant's mental model accuracy because it provided the highest predictor-criterion correlation in comparison to averaged networks with similar coherence scores and number of links.

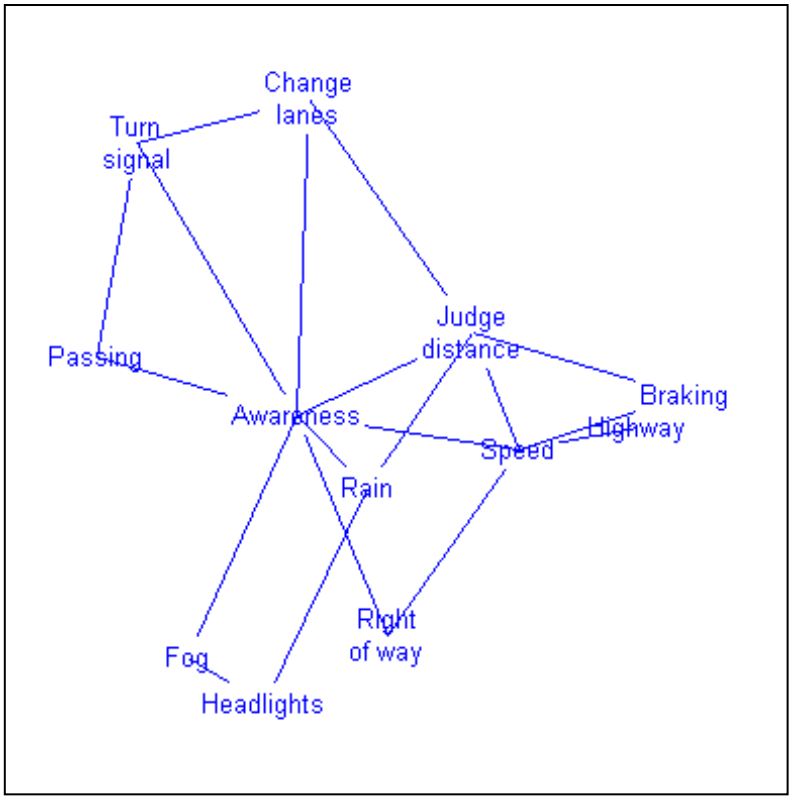

Figure 2. Mental model based on SME2 and SME5 (19 links, *C* = .40)

5.3 Mental models reliability estimates

Participants' mean coherence score was .20 (*SD* = .25) at Administration 1, and .19 (*SD* = .29) at Administration 2. The correlation between mental model coherence scores at Administration 1 and 2 was positive and statistically significant, $r = .31$, $p <$ .01. At the same time, using the .20 cut-off score, 59% of the participants displayed equally coherent (or equally incoherent) mental models between administrations.

Mental model accuracy score correlations, within participant correlation, and within participant similarity were calculated as a means of assessing mental model measurement stability. The correlation between mental model accuracy scores at

Administration 1 and Administration 2 was .49, $p < .001$. Mean within participant correlation was .44 ($SD = .22$) whereas mean within participant similarity (i.e., the degree of similarity between an individual's mental models obtained in two occasions) was .24 ($SD = .15$). Pathfinder offers a way for assessing mental model similarity using a *t*-test in which *p* values lower than .05 indicate that two networks share more common links than what would be expected by chance. The later analysis showed that 82.3% of the sample generated similar mental models between administrations.

Although the within participant correlation was low relative to test-retest reliability standards, within participant similarity considered as a dichotomous variable suggests the opposite (82.3% of the respondents provided consistent mental models). A similar pattern of results were obtained for the coherence scores; while the correlation between coherence scores measured at Administration 1 and 2 was relatively low ($r = .31$, $p <. 01$), 59% of the participants displayed equally coherent (or equally incoherent) mental models between administrations. These results suggest that mental model reliability estimates may depend on the conceptualization of the underlying latent construct. Although theory posits that differences in declarative knowledge are more quantitative in nature whereas differences in knowledge organization are more qualitative, there is no reason why knowledge organization (or reorganization) could not occur incrementally in the case of driving (cf. Dayton, Durso, & Shepard, 1990). That is, individuals repeatedly exposed to the same task should form new links between different aspects of the task that should be reflected in increasingly better (or possibly worse) mental models. Consequently, mental model coherence and accuracy scores were

conceptualized as continuous variables—because the acquisition of driving knowledge can be thought of as an incremental rather than an all-or-none learning process—and all analyses are based on this assumption, including the interpretation of the preceding reliability estimates.

5.4 Contrast of main hypotheses of the study

In support of Hypothesis 1b, mental model accuracy measured at Administration 2 was negatively correlated with tickets ($r = -.18$, $p < .05$) which means that lower mental model accuracy was associated with higher rate of moving violations. Moreover, because declarative knowledge—as measured by the driving knowledge test—was not correlated with either criteria, these initial results also support Hypothesis 2b.

Hypotheses 3a and 3b stated that mental model accuracy would predict a significant amount of variance over commonly used predictors of driving outcomes. To test these hypotheses, a series of hierarchical analyses were conducted to compare the contribution of each variable to the prediction of crashes and tickets. First, each variable was entered in the model according to its absolute correlation to at-fault crashes (Table 7); then, the same procedure was followed using moving violations as the criterion (Table 8).

As shown in Table 7, conscientiousness and sex were related to at-fault crashes. Specifically, low conscientious individuals reported having been involved in more crashes than high conscientious individuals, and females had more crashes than males.

.

Table 7
Hierarchical Regression Analysis Predicting At-fault Crashes

Predictor	β	R^2	ΔR^2
1. Conscientiousness	–0.25	.04*	
2. Sex[a]	–0.24	.09**	.04*
3. Miles per week	0.12	.11**	.02
4. Extroversion	0.14	.13**	.02
5. Speed	0.15	.15**	.02
6. Driving years	–0.13	.15**	.01
7. Emotional Stability	–0.02	.15**	.00
8. Age	0.02	.15**	.00
9. Agreeableness	0.12	.16**	.01
10. GMA	–0.09	.17**	.01
11. Specific LoC	0.09	.18**	.01
12. Coherence adm. 2	0.04	.18*	.00
13. Coherence adm. 1	–0.04	.19*	.00
14. Openness	–0.07	.19*	.00
15. Driving test	–0.02	.19*	.00
16. General LoC	0.03	.19*	.00
17. Accuracy adm. 2	0.08	.20	.00

Note. N = 130. GMA = general mental ability; LoC = locus of control; adm. = administration. [a]Dummy codes for sex are female = 0, and male = 1. *p < .05. **p < .01.

With moving violations as the criterion, miles driven per week, conscientiousness, and sex predicted a significant amount of variance in number of tickets. Therefore, the results presented in Tables 7 and 8 do not support either Hypothesis 3a or Hypothesis 3b. Specifically, no evidence of incremental validity for mental model accuracy was found when assessed in conjunction with conscientiousness and sex for predicting at-fault crashes, or when assessed in conjunction with miles driven per week, conscientiousness, and sex for predicting moving violations.

Table 8
Hierarchical Regression Analysis Predicting Moving Violations

Predictor	β	R^2	ΔR^2
1. Miles per week	.27	.12***	
2. Conscientiousness	−.18	.15***	.03**
3. Driving years	.08	.17***	.02
4. Age	.11	.18***	.00
5. Accuracy adm. 2	−.14	.20***	.02
6. Openness	−.13	.21***	.01
7. Accuracy adm. 1	.01	.21***	.00
8. Sex[a]	−.17	.24***	.03*
9. Emotional Stability	.02	.24***	.00
10. Extroversion	.12	.25***	.01
11. Coherence adm. 2	−.03	.25***	.00
12. Driving test	.00	.25***	.00
13. General LoC	.02	.25***	.00
14. Specific LoC	.05	.25***	.00
15. Agreeableness	−.07	.26***	.00
16. GMA	.04	.26***	.00
17. Coherence adm. 1	.04	.26***	.00

Note. N = 130. GMA = general mental ability; LoC = locus of control; adm. = administration. [a]Dummy codes for sex are female = 0, and male = 1. *p < .05. **p < .01. ***p < .001.

To further investigate this Hypotheses 3a and 3b, a series of regressions were used to compare the effect of mental model accuracy over and beyond individual predictors of driving outcomes, specifically against those predictors that demonstrated statistically significant correlations with driving outcomes in this study sample (Table 9). These analyses showed that mental model accuracy predicted a significant amount of variance in moving violations after partialling out the effects of miles driven per week, and conscientiousness. (Note that ΔR^2 for predicting moving violations after controlling

for sex was not interpreted because sex was not correlated with moving violations.) No evidence of incremental variance in predicting at-fault crashes was found.

Table 9
Mental Model Accuracy Effectiveness for Predicting Crashes and Moving Violations Over and Above Individual Predictors of Driving Outcomes

	At-fault crashes		Moving violations	
Predictor	R^2	ΔR^2	R^2	ΔR^2
Model 1				
Sex[a]	.03*		.01	
Accuracy adm. 2	.03	.00	.04	.03*
Model 2				
Age	.00		.04*	
Accuracy adm. 2	.01	.00	.07*	.03
Model 3				
Driving years	.02		.05*	
Accuracy adm. 2	.02	.00	.07*	.02
Model 4				
Miles per week	.03		.12***	
Accuracy adm. 2	.03	.00	.15***	.03*
Model 5				
Conscientiousness	.04*		.06***	
Accuracy adm. 2	.05	.00	.08***	.03*

Note. $N = 130$. GMA = general mental ability; LoC = locus of control; adm. = administration. [a]Dummy codes for sex are female = 0, and male = 1. *$p < .05$. **$p < .01$. **$p < .001$.

It was hypothesized that driving experience and personality variables typically associated with driving outcomes would correlate with mental model accuracy (Hypotheses 4 and 5, respectively). The correlations between driving years and mental model accuracy at Administration 1 and 2 were not statistically significant (Table 3). In contrast, conscientiousness and emotional stability were positively correlated with

mental model accuracy at Administration 1, $r = .23$, $p < .01$, and $r = .28$, $p < .01$. However, only emotional stability was also correlated with mental model accuracy at Administration 2, $r = .21$, $p < .05$. Finally, in support of Hypothesis 6, GMA was positively correlated with mental model accuracy at Administration 1, $r = .31$, $p < .001$, and Administration 2 ($r = .24$, $p < .001$).

To gain a better understanding of the relationship between driving mental models and other predictors of driving outcomes, a hierarchical regression analysis was conducted using mental model accuracy as the criterion. Results of these analyses are displayed in Table 10. The best predictor of mental model accuracy was declarative knowledge as measured by the driving knowledge test; however this correlation was significant for Administration 2 but not for Administration 1. After controlling for declarative knowledge, GMA and driving years were the best predictors of mental model accuracy. Although emotional stability was positively correlated with mental model accuracy at both Administration 1 and 2, this correlation was no longer statistically significant after controlling simultaneously for GMA and driving years.

Table 10
Hierarchical Regression Predicting Mental Model Accuracy From Other Predictors of Driving Outcomes

Predictor	β	R^2	ΔR^2
1. Driving test	.25	.06**	
2. GMA	.18	.10**	.04*
3. Emotional Stability	.14	.12***	.02
4. Driving years	−.22	.16***	.04*
5. Specific LoC	−.11	.18***	.01
6. Age	.04	.18***	.00
7. Openness	.01	.18**	.00
8. Agreeableness	−.12	.18**	.01
9. General LoC	−.01	.18**	.00
10. Conscientiousness	.00	.18**	.00
11. Miles driven per week	−.08	.19**	.00
12. Extroversion	.06	.19*	.00
13. Sex[a]	−.04	.19*	.00

Note. N = 130. GMA = general mental ability; LoC = locus of control. [a]Dummy codes for sex are female = 0, and male = 1. *p < .05. **p < .01. **p < .001.

However, when controlling for GMA and driving years separately and individually, the association between emotional stability and mental model accuracy was positive and statistically significant (Table 11). In other words, high emotional stability individuals provided more accurate mental models.

Table 11
Emotional Stability as a Predictor of Mental Model Accuracy Over GMA and Declarative Knowledge

	Accuracy Adm. 2	
Predictor	R^2	ΔR^2
Model 1		
GMA	.06**	
Emotional stability	.09**	.03*
Model 2		
Driving knowledge test	.06**	
Emotional stability	.09*	.03*

Note. $N = 130$. GMA = general mental ability; adm. = administration. *$p < .05$. **$p < .01$. **$p < .001$.

6. DISCUSSION

Several predictors were used to predict crashes and moving violations. Consistent with prior research in this domain, demographic, exposure variables, and conscientiousness, were good individual predictors of crashes and moving violations. Other associations between personality variables typically used in this domain—such as agreeableness, and extraversion (cf. Clarke & Robertson, 2005)—were not correlated with criteria.

As hypothesized, mental model accuracy was better at predicting driving outcomes in comparison to a traditional driving knowledge test. This study also provided support for mental models as a predictor of driving outcomes over and beyond the effect of a few statistically significant predictors (i.e., miles driven per week and conscientiousness). Given that conscientiousness and mental models were the only statistically significant predictors of driving outcomes from the psychological domain, the generalizability of these findings may be questioned. For instance, previous studies have found that agreeableness and extroversion are associated with driving outcomes (Clarke & Robertson, 2005); however, the validity of these predictors was not replicated in the context of this study. Thus, it would be improper to conclude that, in general, mental models add incremental validity to previously established predictors of driving outcomes. The problem with replicating some previous findings may be due to the current study's limited sample. Consequently, using a more representative sample of the

population of drivers is strongly recommended in future studies (e.g., this sample's age ranged from 17 to 21 years).

Contrary to Hypotheses 4 and 5 expectations, neither driving years nor personality variables typically associated with driving outcomes were correlated with mental model accuracy. One exception was emotional stability, which was correlated with mental model accuracy at both Administration 1 and 2. Hansen (1989) pointed to a mechanism linking emotional stability and crashes via attention to the task. This hypothesis states that individuals low in emotional stability may be more sensitive to stressors which, in turn, may divert their attention and deplete cognitive resources for completing the task at hand. Therefore, although emotional stability did not predict crashes or moving violations, the distractibility hypothesis would suggest that low emotionally stable individuals would be more prone to make errors when completing a task that is cognitively demanding, such as the mental model production task that was used for collecting driving mental models. (As it will be argued later, several factors including GMA may influence mental model measurement reliability.)

Some methodological issues about the measurement of mental models and the choice of the most appropriate referent structure were also addressed. First, mean within participant similarity (M_C = .24, SD = .15) and mean within participant mental model correlations (M_r = .44, SD = .22) were relatively low, which indicates low test-retest reliability. Also, whereas mean coherence scores were almost the same for Administration 1 (M_{coh} = .20, SD = .25) and Administration 2 (M_{coh} = .19, SD = .29), the correlation between coherence scores between administrations was only moderate (r =

.31, $p < .001$). These results indicate that mental model measures were only moderately reliable, and further research is needed in order to understand if this is a sample specific issue or a larger issue within mental models that may be causing this phenomenon. Notwithstanding, since validity is typically reduced by predictor unreliability, it is plausible that the correlation between mental models and driving outcomes should improve along with more reliable mental model assessments.

There were statistically significant correlations between within participant mental model similarity and GMA, and to driving knowledge test scores. The fact that GMA was associated with an individual's capacity to provide consistent mental models, may be due in part to the novelty of this task (it is unusual for individuals to complete pairwise comparisons). If this is the case, researchers should pay special attention to the instructions they give participants to complete this task. This may involve such actions as simply asking them if they have fully understood the instructions, or providing practice items and checking the accuracy of their responses *before* proceeding with the actual test.

There is yet another recommendation that stems from the way in which the task is presented. For this study each focal concept was presented at the center of a bull's eye type diagram and participants were instructed to drag-and-drop the remaining concepts around this target. Although several terms can be arranged around the focal concept at once, Pathfinder will only record the distance between a term and the focal concept and will not record information about the relatedness between terms other than the focal concept. Therefore, if a participant mistakenly arranges the terms based on their

relatedness with the target *and* between each other he/she might provide information different that the information he/she would have provided if only pairwise comparisons were presented. If this is the case, error variance due to the method used for collecting pairwise comparison scores may influence mental model reliability, validity, or both. At the same time, thinking of these concepts simultaneously may be more challenging than thinking about their relatedness one at a time. Using a simpler task for collecting pairwise comparisons (e.g., a slide bar between concepts) may reduce the impact of GMA on mental model reliability.

Driving knowledge test scores were positively correlated with mental model accuracy ($r = .25$, $p < .01$). This result indicates that mental models, although associated with declarative knowledge, are substantially different from it. Also, because declarative knowledge was positively correlated with within participant mental model similarity ($r = .23$, $p < .01$) mental model reliability may also depend on an individual's previous knowledge of the domain. In other words, variance in declarative knowledge may be associated with variance in mental model reliability, such that individuals deficient in declarative knowledge may experience difficulties in providing consistent mental models. This idea is consistent with Anderson's (1982) stages of skill acquisition, wherein declarative knowledge precedes procedural knowledge—which is the aspect of knowledge that mental models are purportedly measuring.

Selecting terms for generating mental models is undoubtedly a critical aspect of mental model measurement. For this study a literature-review-based approach was implemented for inferring what was the most appropriate set of terms for this measure.

Basically, the themes extracted from the driving education textbook were translated into one word or short phrase to represent the content domain. Although this procedure may not be optimal, all the terms were considered important by the SMEs.[4] Needless to say, the ability of the researchers to determine what terms are more important was partly due to their own driving experience. For other tasks (e.g., landing a large commercial jetliner) it would have been impossible to come up with equally relevant terms. As of now, a second study is being conducted in which a number of transportation experts are providing input on what categories of tasks are most relevant for predicting driving outcomes. This later approach (i.e., in-depth task analysis using SMEs) seems more adequate for inferring what the most appropriate set of mental model terms is, although the superiority of this process (e.g., for improving criterion-related validity) is yet to be empirically demonstrated.

It was previously mentioned that coherence scores are not necessarily good indicators of mental model internal consistency. In fact, like accuracy scores, it is argued that coherence scores should be better conceptualized as an index of an individual's mastery of the knowledge domain. Finally, because both accuracy scores and coherence scores displayed higher correlations with declarative knowledge for Administration 2, it is possible that coherence scores may be also influenced by an individual's ability to understand the pairwise comparison task.

A related issue is the lack of explicit guidelines for choosing the best referent structure to derive mental model accuracy scores. For the purpose of the present study,

[4] SMEs were required to rate the importance of the terms using a scale from 1 to 5. The mean importance rating for the terms ranged between 3.20 to 5.00.

the best referent structure was chosen because of its high correlation with criteria; nonetheless, this referent structure was different from the one that would have been chosen based on mental model properties such as coherence scores and number of links. For instance, the best referent structure in terms of coherence score (C = .51) and number of links (N = 19) was SME34 (see Tables 4 and 5). The second best referent mental model was the one that was actually used for deriving accuracy scores, namely, SME25. Interestingly, an averaged network of SMEs 2, 3, 4, and 5, did not improve the criterion-related validity of mental model accuracy scores. In sum, the best or "ideal" referent structure was not the best predictive referent mental model for the purposes of this study.

To our knowledge, there are no clear guidelines on how to combine expert mental models to derive one single referent structure. This issue is particularly important when (a) experts' mental models are uncorrelated, and (b) when combinations of expert mental models yield varying, yet acceptable, coherence scores. On one hand, if expert mental models are perfectly correlated, any one can be used as a referent structure separately and individually. On the other hand, if they are uncorrelated, a combination of them may decrease their individual validity. Results of the present study suggest that a combination of expert mental models may capture the criterion space more accurately, with each independent expert contribution improving our understanding of the domain. In fact, when adding SMEs, the criterion-related validity of mental models also increased (although at the expense of mental model coherence). Needless to say, deciding who is an expert and why is also critical. As in job analysis, past performance

and/or reputation may be used for selecting SMEs, but their ability to convey relevant and articulated information about what they do, should also be taken into account.

7. CONCLUSION

Driving mental models' incremental validity was supported, although the variance explained by mental model accuracy was relatively small and was moderated by the criteria used to operationalize driving outcomes (i.e., at-fault crashes versus moving violations). Declarative knowledge and GMA were positively correlated with mental model accuracy, which indicates that they may play an important role as antecedents of mental model development. One of the Big Five personality factors consistently associated with mental model accuracy was emotional stability—such that high emotionally stable participants provided more accurate mental models at Administration 1 and Administration 2.

The temporal stability of mental models was low either when assessed using within participants mental model accuracy ($r = .49$, $p < .001$), within participants mental model similarity ($M_C = .24$, $SD = .15$) or within participant mental model correlations ($M_r = .44$, $SD = .22$). The low test-retest reliability of this measure may be due to difficulties inherent to mental model measurement procedure. Two recommendations for reducing unreliable mental models variance are (a) to provide clear instructions to participants as well as practice items before taking the test; and (b) to reduce the complexity of the task by presenting isolated pairs of items instead of the bull's eye diagram in which all the concepts are presented simultaneously.

It is argued that previous experience of test developers with the knowledge domain should not replace expert judgments, and steps should be taken to ensure that the

knowledge domain is sufficiently and accurately represented. Selecting mental model terms and identifying an appropriate referent structure for deriving mental model accuracy scores are critical steps, and further research is needed to address these issues.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology, 117*, 288-318.

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist, 38*, 85-93.

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science, 16*, 235-239.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219-245.

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology, 86*, 303-311.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates,.

Arthur, W. Jr., Barret, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance, 4*, 89-105.

Arthur, W. Jr., Barrett, G. V., & Doverspike, D. (1990). Validation of an information-processing-based test battery for the prediction of handling accidents among

petroleum-product transport drivers. *Journal of Applied Psychology, 75*, 621-628.

Arthur, W. Jr., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement, 54*, 394-403.

Arthur, W. Jr., & Doverspike, D. (1992). Locus of control and auditory selective attention as predictors of driving accident involvement: A comparative longitudinal investigation. *Journal of Safety Research, 23*, 73-80.

Arthur, W. Jr., & Doverspike, D. (2001). Predicting motor vehicle crash involvement from a personality measure and a driving knowledge test. *Journal of Prevention & Intervention in the Community, 22*, 35-42.

Arthur, W. Jr., Fuentes, R., & Doverspike, D. (1990). Relationships among personnel tests, age, and job performance. *Experimental Aging Research, 16*, 11-16.

Arthur, W. Jr., & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality, 64*, 593-618.

Arthur, W. Jr., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven Advanced Progressive Matrices Test. *Journal of Psychoeducational Assessment, 17*, 354-361.

Arthur, W. Jr., & Day, E. A. (2009). Information processing, personality, and demographic variables as predictors of crashes and moving violations. In H.

Bédard & G. Delashmit (Eds.), *Accidents: Causes, analysis, and prevention* (pp. 125-148). New York: Nova Science Publishers.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Journal of Applied Psychology, 78*, 111-118.

Bellet, T., Bailly-Asuni, B., Mayenobe, P., & Banet, A. (2009). A theoretical and methodological framework for studying and modelling drivers' mental representations. *Safety Science, 47*, 1205-1221.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*, 887-919.

Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78*, 355-376.

Davis, F. D., & Yi, M. Y. (2004). Improving computer skill training: Behavior modeling, symbolic mental rehearsal, and the role of knowledge structures. *Journal of Applied Psychology, 89*, 509-523.

Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: Validity and Utility of structural knowledge assessment. *Journal of Business and Psychology, 18*, 191-206.

Day, E. A., Arthur, W. Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*, 1022-1033.

Dayton, T., Durso, F. T., & Shepard, J. D. (1990). A measure of knowledge reorganization underlying insight. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 267-277). Norwood, NJ: Ablex.

Edwards, B. D., Day, E. A., Arthur, W. Jr., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology, 91*, 727-736.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96.

Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 241-254). New Jersey: Ablex.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83*, 88-96.

Gulliver, P., & Begg, D. (2007). Personality factors as predictors of persistent risky driving behavior and crash involvement among young adults. *Injury Prevention, 13*, 376-381.

Hansen, C. P. (1989). A causal model of the relationship among accidents, biodata, personality, and cognitive factors. *Journal of Applied Psychology, 74*, 81-90.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*, 196-205.

Johnson-Laird, P. N. (1983). *Mental models* (6th ed.). Cambridge, MA: Harvard University Press.

Jonah, B. A. (1997). Sensation seeking and risky driving. In T. Rothengatter & E. Carbonell (Eds.), *Traffic and transport psychology* (pp. 259-267). Oxford: Pergamon.

Koelega, H. S. (1992). Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological Bulletin, 112*, 239-258.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Kraiger, K., Salas, E., & Cannon-Bowers, J. A. (1995). Measuring knowledge organization as a method for assessing learning during training. *Human Factors, 37*, 804-816.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268-279.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points.". *Annual Review of Psychology, 51*, 405-444.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior, 26*, 37-56.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, *36*, 876-910.

Montag, I., & Comrey, A. L. (1987). Internality and externality as correlates of involvement in fatal driving accidents. *Journal of Applied Psychology, 72*, 339-343.

National Highway Traffic Safety Administration (NHTSA), National Center for Statistics and Analysis. (2009a). *2008 Traffic safety annual assessment highlights* (DOT HS 811 172). Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/811172.pdf.

National Highway Traffic Safety Administration (NHTSA), National Center for Statistics and Analysis. (2009b). *Traffic safety facts: 2008 Data* (DOT HS 811 162). Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/811162.pdf.

National Highway Traffic Safety Administration (NHTSA), National Center for Statistics and Analysis. (2009c). *Driver electronic device use in 2008* (DOT HS 811 184). Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/811184.pdf.

Qiang, W., Bowling, N. A., & Eschleman, K. J. (2010). A meta-analytic examination of work and general locus of control. *Journal of Applied Psychology, 95*, 761-768.

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the united States. *Journal of Educational Measurement, 26*, 1-16.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control reinforcement. *Psychological Monographs: General and Applied, 80* (1, Whole No. 609).

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*, 349-363.

Schmidt, A., Beauducel, A., Brocke, B., & Strobel, A. (2004). Vigilance performance and extraversion reconsidered: Some performance differences can indeed be induced. *Personality & Individual Differences, 36*, 1343-1351.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.

Smith-Jentsch, K. A., Campbell, G. E., Milanovich, D. M., & Reynolds, A. M. (2001). Measuring teamwork mental models to support training needs assessment, development, and evaluation: Two empirical studies. *Journal of Organizational Behavior, 22*, 179-194.

Solis, H. L., & Hall, K. (2009). *Fatal workplace injuries in 2006: A collection of data and analysis* (CFOI Report 1015). Retrieved from U. S. Bureau of Labor Statistics, The Census of Fatal Occupational Injuries website: http://www.bls.gov/iif/cfoibulletin2006.htm.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

West, R. J., Elander, J., & French, D. (1993). Mild social deviance, Type-A behaviour pattern and decision-making style as predictors of self-reported driving style and traffic accident risk. *British Journal of Psychology, 84*, 207-219.

Williams, A. F., & Carsten, O. (1989). Driver age and crash involvement. *American Journal of Public Health, 79*, 326-327.

VITA

Name: Gonzalo Javier Muñoz Gálvez

Address: Department of Psychology
Texas A&M University
College Station, TX 77843-4235

Email Address: gmunoz@tamu.edu

Education: M. S., Psychology, Texas A&M University, 2011.
B. S., Psychology, Pontificia Universidad Católica de Chile, 2002.