11-14-2016 12:00 AM

Fitness-to-Drive Screening Measure© : Constructing and Validating the short form

Shabnam Medhizadah, *The University of Western Ontario*

Supervisor: Dr. Sherrilene Classen, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Health and Rehabilitation Sciences

Abstract

The valid and reliable Fitness-to-Drive Screening Measure© (FTDS) is a free online proxy tool that screens for at-risk older drivers using 54 driving-related items. This secondary data analysis aims to construct and validate a short form FTDS, using 200 caregiver FTDS response sets and 200 older driver on-road assessment pass/fail outcomes. To construct the short form, exploratory factor analysis and classical test theory techniques were employed to determine the most interpretable factor model and the minimum number of items that might be used to predict fitness to drive. Next, receiver operating characteristics curves were generated to evaluate the concurrent criterion validity of the constructed short form against driver on-road pass/fail outcomes. This study resulted in the construction of FTDS short form version 1 and 2. Both short forms were predictive of on-road pass/fail outcomes. This measure may provide proxy raters with efficient means to help identify at-risk older drivers.

Keywords

Older drivers, caregivers, exploratory factor analysis, item analysis, receiver operator characteristic (ROC) curve, short form, concurrent criterion validity, screening tool.

Co-Authorship Statement

Principle Author:

Shabnam Medhizadah

Ms. Medhizadah performed all of the analysis, interpreted data, wrote the manuscripts and will act as the corresponding author for the manuscripts included in this thesis document.

Co-Authors:

1. Dr. Sherrilene Classen

Dr. Classen supervised the development of the work, helped with data interpretation and manuscript evaluation.

2. Dr. Andrew Johnson

Dr. Johnson guided the data analysis and editing of the manuscripts.

Dedication

To Mom, Dad, Fahim, Roobina, and Gitie, without you this would not have been possible.

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor Dr. Sherrilene Classen for her continuous support, patience, words of wisdom and belief in me throughout my entire master's degree. Her guidance and mentorship helped me in conducting and writing this thesis.

I would like to also thank Dr. Andrew Johnson for his insightful comments, help and knowledge of statistics.

Last but not least, I would like to thank the i-Mobile Research lab; Dr. Liliana Alvarez, Sarah Krasniuk and Melissa Knott for all of their support and words of encouragement, it would have been a tough journey without you.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AUC	Area under the curve
CDRS	Certified Driver Rehabilitation Specialist
EFA	Exploratory factor analysis
FTDS	Fitness-to-Drive Screening Measure©
FTDS V1	Fitness-to-Drive Screening Measure© short form version 1
FTDS V2	Fitness-to-Drive Screening Measure© short form version 2
GRS	Global Rating Score
KMO	Kaiser-Meyer-Olkin
MVC	Motor vehicle collision
NLR	Negative Likelihood Ratio
NPV	Negative Predictive Value
OT-CDRS	Occupational Therapy - Certified Driver Rehabilitation Specialist
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic curve
U.S.	United States of America
UF	University of Florida
UWO	University of Western Ontario

Chapter 1

1 Introduction

As the number of older adults (64 years of age or older) continues to increase in Western societies, the need to evaluate older drivers fitness to drive is growing. Normal age-related declines in visual, cognitive, and physical functioning may impact an older adult's fitness to drive, defined by Transportation Research Circular (2016, p.10) as the "absence of any functional (sensory-perceptual, cognitive or psychomotor) deficit or medical condition that significantly impairs an individual's ability to fully control the vehicle while conforming to the rules of the road and obeying traffic laws, or that significantly increases crash risk." In fact, older adults are often overrepresented in motor-vehicle collisions, comprising more than 25% of road-traffic injuries in Canada and the United States of America (U.S.) (National Highway Traffic Safety Administration [NHTSA], 2013; Nicolleta, 2013).

Driving is an important occupation for older drivers in Canada and the U.S., who engage in driving throughout most of their lives. As such, driving cessation often has negative social and psychological implications for older drivers, including a negative impact on their perceived quality of life, sense of autonomy and wellbeing (Dickerson, Meuel, Ridenour, & Cooper, 2014; O'Neill, Bruce, Kirby, & Lawlor, 2000). Thus, it is crucial to start timely conversations about when drivers should seek professional help allowing them to stay on the road safely for longer, or to consider driving cessation and other community mobility options in a timely manner. As such, using screening tools like the Fitness-to-Drive Screening Measure© (FTDS) may assist older drivers in starting positive and timely conversations about their fitness to drive abilities.

The FTDS is an online screening measure that proxy raters (formal/informal caregivers, family members or friends) can use to identify at-risk older drivers. This free, low-risk measure takes approximately 20 minutes to complete and consists of 54 driving-related items. Empirical support suggests that the FTDS is a valid and reliable measure for screening at-risk older drivers (Classen, Velozo, Winter, Bédard, & Wang, 2015). However, despite being accessed by more than 22,000 users across Canada and the U.S., a recent report by Classen, Medhizadah and Alvarez (2016) examining FTDS user patterns and trends found that many users were not completing the measure. For instance, many proxy raters "left" the FTDS website from the demographics webpage, where they were asked to provide information about themselves, potentially because of perceived irrelevance to the task at hand (Classen, Medhizadah, & Alvarez, 2016). Likewise, many users "left" the FTDS website from the end user agreement webpage, where they were asked to accept the terms of use before proceeding to the FTDS. This may have been due to too much text on the webpage or users not agreeing with the terms of use (Classen et al., 2016). The report also indicated that many users were not completing the FTDS in its entirety potentially because of the measure's required completion time. Classen and colleagues (2016) suggested the development of a short form FTDS as a means of potentially tackling this issue. This study addresses the suggestion to develop a short form FTDS, that may decrease the completion time of the measure and potentially increase user completion rates. Thus the aim of this thesis project is to construct and validate a short form FTDS.

1.1 References

Classen, S., Medhizadah, S., & Alvarez, L. (2016). The Fitness-to-Drive Screening Measure©: Patterns and trends for Canadian users. *The Open Journal of Occupational Therapy, 4(4),* 1-15. doi.10.15453/2168-6408.1227

Classen, S., Velozo, C. A., Winter, S. M., Bédard, M., & Wang, Y. (2015). Psychometrics of the Fitness-to-Drive Screening Measure©. *OTJR: Occupation, Participation and Health, 35*(1), 42-52. doi.10.1177/1539449214561761

Dickerson, A. E., Meuel, D. B., Ridenour, C. D., & Cooper, K. (2014). Assessment tools predicting fitness to drive in older adults: A systematic review. *The American Journal of Occupational Therapy, 68*(6), 670-680. doi: 10.5014/ajot.2014.011833

National Highway Traffic Safety Administration. (2013). Traffic safety: data 2013. U.S. National Department of National Transportation. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812169

Nicoletta, N. (2013). Driving characteristics of young and aging population. Statistics Canada.

O'Neill, D., Bruce, I., Kirby, M., Lawlor, B. (2000) Older drivers, driving practices and health issues. *Clinical Gerontologist, 22*, 47-54.

Transportation Research Circular. (2016). *A taxonomy and terms for stakeholders in senior mobility*. Transportation Research Circular No.E-C211. Washington, DC: Transportation Research Board.

Chapter 2

2 Constructing and Validating the Short Form Fitness-to-Drive Screening Measure©

2.1 Background

In Canada and the United States of America (U.S.), older adults (65 years of age or older) are overrepresented in fatal motor vehicle collisions (MVCs). In fact, this population comprises 24.5% of Canadian and 27.0% of American road traffic-related deaths and injuries (National Highway Traffic Safety Administration [NHTSA], 2013; Nicolleta, 2013). The high number of fatal MVCs amongst this population may be partially attributed to increasing frailty due to ageing (Eby, Molnar, Shope, Vivoda, & Fordyce, 2003; NHTSA, 2013). Normal age-related declines in physical, cognitive, and neurological functions are inevitable and may affect a person's fitness to drive (i.e., the physical and mental abilities and resources required for driving a motor vehicle without impeding the progress of other road users) (Brouwer & Ponds, 1994). Currently, more than 54 million older adults are licensed drivers in Canada and the U.S., with numbers projected to increase due to the growth of this population (Federal Highway Administration, 2014; Nicolleta, 2013).

Mirroring this projection is the growing need to support older drivers and their caregivers, by recognizing the driver's changing abilities and adapting their driving practices to stay on the road safely for as long as possible. The Fitness-to-Drive Screening Measure© (FTDS), a web-based measure developed to help caregivers identify at-risk older drivers, may help provide such support. The FTDS equips caregivers with strategies and resources for decision-making related to driving, referral to rehabilitation, or driving cessation for the driver. Although the measure has been used quite extensively, a recent study (Classen, Medhizadah, & Alvarez, 2016) revealed that a short form FTDS (vs. the existing 54-item version) may potentially improve its use and impact. As such this study's overarching objective was to examine if a short form FTDS could be developed.

2.2 Literature Review

Overall, older adults overestimate their fitness to drive abilities when conducting self-evaluations and as such demonstrate a lack of awareness of their driving abilities (Marottoli & Richardson, 1998). Lack of awareness, defined by Marotolli and Richardson (1998), as the "discrepancy between one's perception of ability and actual ability," (p.332) is critical for identifying one's driving limitations and adapting accordingly. Rather than having older drivers provide biased self-evaluations of their fitness to drive abilities, their caregivers (i.e., friends, family) who have lived experiences with them, can provide proxy assessments with more accurate reflections. For example, Classen, Velozo, Winter, Bédard, and Wang (2015) reported that caregiver evaluations of older drivers' abilities were more consistent and reliable than the older drivers' self-evaluations.

Most often, caregivers are the primary source for providing assistance with activities of daily living to older adults (Alecxih, Zeruld, & Olearczyk, 2001). In Canada and the U.S., caregivers are mostly female making up between 59-70% of the caregiver population for older adults (Alecxih et al., 2001; Family Caregiver Alliance, 2010; Hollander, Liu, & Chappel, 2009). Female caregivers are usually between the ages of 34-64 years old and 38% of them are the spouse, friend or neighbor of the older adult

(Alecxih et al., 2001; Family Caregiver Alliance, 2010). These caregivers spend an average of 24 hours a week on caregiving duties with the number of such hours increasing with the age of the caregiver (Family Caregiver Alliance, 2010; National Alliance for Caregiving and AARP, 2015). Due to their extensive involvement, caregivers can act as primary prevention agents for identifying a change in driving abilities, and for suggesting adaptive actions toward older drivers. One such strategy may include initiating conversations about driving cessation (McPeek, Nichols, Classen, & Breiner, 2011; Musselwhite & Shergold, 2013). Although older drivers report to be somewhat upset by caregivers' advice to cease driving, 60% of them follow such counsel (Musselwhite & Shergold, 2013). Therefore, caregivers are a valuable resource for increasing awareness among older drivers and their fitness to drive abilities.

Although driving cessation may be inevitable for some older drivers, older drivers and their caregivers may benefit from having access to tools to further assist them with driving-related decisions. Specifically, through tools designed to assist caregivers, such as the Fitness-to-Drive Screening Measure© (FTDS), caregivers may be informed of drivers' abilities, and receive resources and recommendations to either prolong the older driver's ability to stay on the road safer for a longer time, or to consider steps toward driving cessation.

The FTDS is a free web-based screening measure available for proxy raters (e.g., formal/informal caregivers, family members or friends). The goal of the FTDS is to empower caregivers to help identify at-risk older drivers. Canadian and American versions of the FTDS are available at www.fitnesstodrivescreening.com. Taking approximately 20 minutes to complete, this tool is comprised of three sections:

- Section A includes demographic questions about the driver and the rater;
- Section B contains questions about the driver's history;
- Section C covers 54 driving-related items geared towards finding information on the drivers' driving behaviors. This section requires the rater to use observations, of the older driver, made in the last three months. Specifically, caregivers' need to rate items progressing from easy, such as "How difficult is it for the driver to drive in the proper lane?" to more challenging, such as "How difficult is it for the driver to drive when there is fog". The items are rated on a four-point scale, ranging from *not difficult* to *very difficult.* Appendix A contains the FTDS questionnaire.

Once raters complete these sections, they are provided with:

(1) A classification of the driver as one of three groups, which may be:

- *At-risk driver*: there are immediate safety concerns that must be addressed immediately;
- *Routine driver*: there are some safety concerns presents, early signs of needing intervention exist; and
- *Accomplished driver*: there are no safety concerns present, however, the driver may experience difficulty with some challenging driving situations.

(2) Specific resources and recommendations according to the classification. For example, at-risk drivers will be advised to arrange for a comprehensive driving evaluation with a Certified Driver Rehabilitation Specialist (CDRS), and receive complimenting resources

such as the link to the Association for Driver Rehabilitation Specialists webpage for finding a CDRS.

(3) A key form summarizing the results of section C, which highlights the driver's overall difficulty based on the caregiver's item ratings. Based on the key form results and the accompanying information obtained from the resources and recommendations, raters may plan and initiate further steps for the driver. For example, raters may use the information provided to raise awareness and inform healthcare professionals about the challenges their loved one face when driving.

Since the FTDS became available online, 22,849 new users have accessed it across Canada and the U.S. (Google Inc., 2014). Despite the high number of new users, people may not be necessarily completing the screening measure in its entirety. We recently explored the Canadian user patterns of the FTDS and found that users did not spend the recommended 20 minutes to complete the FTDS rating on the website. Also, 60% of users left the website before reaching the final results or key form webpages (Classen et al., 2016).

Thus, currently, the FTDS is being accessed and used by more than 22,000 people. However, to increase its impact and reach, a short form FTDS is recommended (Classen et al., 2016). A short form may reduce the completion time of the screening measure, and potentially ensure that more users will complete the FTDS as intended. Moreover, the literature indicates that shorter (vs. longer) questionnaires result in higher user responses (Rolstad, Adler, & Ryden, 2011; Widaman, Little, Preacher, & Sawalani, 2008).

Psychometrics of the FTDS

The FTDS is a valid and reliable screening tool for identifying at-risk older drivers (Classen, Velozo, Winter, Bédard, & Wang, 2015). Exploratory factor analysis (EFA) of the measure suggested that a 2-factor model best represented the constructs within the FTDS. However, after removing 14 “pre-driving items” (e.g., “Can she or he get in the car?”), the remaining items contributed more homogenously to a one-factor model representing the fitness to drive construct. Confirmatory factor analysis suggested that evaluator and caregiver ratings fit a one-factor model, whereas the driver ratings did not. Specifically, evaluator and caregiver ratings indicated good unidimensionality by meeting the criteria for the comparative fit index, and Tucker-Lewis index, but not the root mean square of error approximations criteria (Classen et al., 2015).

Principles of Rasch analysis were used to examine item and person level properties. The Rasch measurement model is often used to validate scale construction because of its (1) ability to determine the relative difficulty of an item in a scale, sequentially ranging from easy to difficult, and (2) ability to provide an estimate of a person’s standing relative to the measure (Bond & Fox, 2001, p.31). For measures with dichotomous rating scales, Rasch analysis converts raw item scores to logits. A logit is a unit on a log odd (interval) scale that has the same value or “distance” between logits (Bond & Fox, 2001, p.17). Next, logit values for the items are mapped along a linear continuum that can range from negative to positive values. The “higher up” the item is on the continuum, the greater the difficulty of the item. Similarly, the placement of a person's logit score on the same linear continuum indicates the individual's ability as contrasted to the difficulty of the item. The greater the person's logit score on the continuum, the higher the probability that the person will be able to complete lower level items successfully.

In Rasch analysis, tools with polytomous rating scales, such as the FTDS, have an additional attribute. That is, for each item on the continuum, there are thresholds that separate each potential response on the rating scale (e.g., Likert scale). The "distance" between thresholds (potential response on the scale) is not considered to be the same or equal. Overall, for polytomous scales each item on the logit scale is considered to have a continuum of its own corresponding to the rating scale. Rasch analysis utilizes user response data to estimate the distance between thresholds of the rating scale for all the items. Once the distance between the thresholds is determined, it is applied to all of the items in the logit scale (Bond & Fox, 2014, p.105).

Examining polytomous rating scales using Rasch analysis provides researchers with information about how the rating scale is utilized. It can reveal rating scale design issues and areas of confusion or misuse by users. Therefore, Classen and colleagues (2015) investigated the FTDS rating scale (5 level adjectival scale: cannot do, very difficult, somewhat difficulty, a little difficult, not difficult) using Rasch analysis. Findings revealed that two rating categories *cannot do,* and *very difficult* were underutilized by drivers, evaluators, and caregivers. Therefore the categories mentioned above were combined into one rating category, *very difficult*. Combining the categories resulted in better utilization of all levels of the rating scale, as evidenced in the study conducted by Classen and colleagues (2015).

Inter-rater reliability on the FTDS among the three groups of raters, indicated significant but weak correlations (ICC = .25, $p < .05$). The strongest correlation (albeit still moderate) emerged between the ratings of driving evaluators and caregivers (ICC = .39, $p < .05$; Classen et al., 2015). Furthermore, as expected, evaluators were more severe in their ratings of drivers than caregivers and driver self-evaluations for 19 of the items, $\chi^2(2) = 586.1$, $p < .05$ (Classen et al., 2015).

Classen et al. (2015) also employed a Receiver Operating Characteristic (ROC) curve, to determine a cutoff score for the FTDS that predicted on-road assessment pass/fail outcomes. Although the area under the curve (AUC) was statistically significant for both caregiver ratings (.73, $p < .05$) and driver ratings (.62, $p < .05$), only the caregiver ratings met the AUC criterion (> .70) set within the study. This suggests that when used by a caregiver, the FTDS can correctly identify an at-risk driver. Classen et al. (2015) identified five cut points with varying sensitivity and specificity values, and these are presented in Table 4.

Gaps

Despite the positive psychometric properties of the FTDS, the tool is not being completed by many of its users, potentially because of the time it takes for completion (about 20 minutes). Classen and colleagues (2016) identified several aspects of the FTDS that if refined by constructing a short form, could potentially increase the tool's uptake and impact. These aspects include (1) decreasing the length of the tool, which may potentially result in more users completing the FTDS and reaching the final recommendations and key form page; and (2) decreasing the overall time it takes to complete the tool, which would potentially increase user response rates. Therefore, by constructing a short form FTDS this study may decrease the time it takes to complete the tool.

Purpose

There are three major impetuses for this study. First, licensed older drivers are increasing in numbers in both Canada and the U.S. Additionally, many older drivers are living beyond their driving lifetime, and due to age-related changes may be more likely to be injured or die when involved in a MVC. Inevitable age-related changes can affect fitness to drive reiterating the need to screen for it. Second, caregivers are in the position to provide valuable information about the fitness to drive abilities of older adults. Moreover, with the use of tools such as the FTDS, the caregiver can identify at-risk older drivers and take the necessary next steps. Third, it is critical that the assessment tool be of an appropriate length and difficulty, in order to facilitate completion by caregivers. By constructing a short form FTDS, this tool may increase the FTDS's completion rate. Thus, the purpose of this study was to construct a reliable and valid short form FTDS.

2.3 Methods

Primary data collection, and subsequent release of the data to researchers at the University of Western Ontario (UWO), was approved by the Institutional Review Board at the University of Florida (UF). The UWO's Office of Human Research Ethics certified this study as a board review exemption (FWA00000121), on March 27, 2015. Appendix B and C contains the ethics letters from the UF and UWO.

Participants

This secondary analysis used de-identified data previously collected by Classen and colleagues (2015) for the Psychometrics of the Fitness-to-Drive Screening Measure© study (hereafter, referred to as the primary study) was used. The data used included 200 caregivers of older adults (family members, friends or formal caregivers), recruited in

North-Central Florida, United States, and Thunder Bay, Ontario, Canada. Participants were between the ages of 18 and 85 years, and were included if they had had the opportunity to observe the older drivers' driving behaviors within the past three months. Participants were excluded if they displayed signs (through telephone interview and observation) of physical or mental conditions that may impair their ability to complete the FTDS, or make valid observations about the older drivers' driving behaviors.

Procedure

The secondary analysis described in this paper followed a four-step process to construct and validate a short form of the FTDS: (1) quality control to ensure data completeness (2) exploratory factor analysis; (3) item analysis; and (4) correlational analysis.

Data Collection and Management

Quality control for the data used in these analyses was performed to ensure data completeness. Items that displayed a pattern of missing responses across participants were removed. For example, item 53 "Drive on a snow covered road" was not answered by more than a quarter of the participants. Upon closer observation it was noted that this specific item was not missing responses at-random. A pattern had emerged where drivers (probably from Florida) were not completing this item because they may have never driven in snow. Identifying similar patterns among other items, a total of nine items were excluded from the analysis, and so the final analysis included 45 of the 54 items on the original measure. The removed items (described in abbreviated form) were:

20. Drive on graded (unpaved) road

26. Parallel park

35. Drive in an unfamiliar urban area

36. Control his or her car when going down a steep hill

37. Exit an expressway or inter-state highway from the left-handed lane

38. Drive in a highly complex situation (such as a large city with high-speed traffic, multiple highway interchanges and several signs)

40. Drive a different vehicle (such as another person's car or a rental car)

53. Drive on a snow covered road

54. Drive on an icy road.

Ten caregiver participants were judged to have insufficient data to be included within the analysis because items were still missing responses, resulting in a final sample size of 190 completed FTDS caregiver response sets for the secondary data analysis.

Data Analysis

Data analysis was done using R software version 3.1.2 (R Core Team, 2015).

Exploratory factor analysis. We employed an exploratory factor analysis to examine the number of underlying factors in the FTDS, within the 45 items that were included in this secondary data analysis. Factorability of the data was assessed using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The number of factors extracted within the analysis was determined through the use of parallel analysis. This method involves comparing the eigenvalues derived within one's analysis against eigenvalues that are derived in a series of factor analyses that are conducted on randomly generated data, that conform to the shape of the matrix used in the empirical analysis (i.e., the randomly generated data will have the same number of participants, and the same number of variables as the empirical data).

Extracted data were rotated to facilitate interpretation, using a Promax rotation, as this method allowed the factors to be intercorrelated. Items that failed to load above 0.40 on any factor were removed from consideration in constructing the short-form of the FTDS, as were items that demonstrated a roughly equivalent dispersion of variability across more than one factor.

Item analysis. To determine the minimum number of items that might be used to predict fitness to drive, we employed a number of classical test theory methods that might suggest an item pool that has adequate reliability and validity. We computed Cronbach's α for each factor, to determine whether the items within the factor might be considered to be measuring the same construct. We then calculated the alpha that would result from the deletion of each individual item (alpha-if-deleted). Items were removed from consideration in the short-form of the FTDS if their deletion from scale produced a substantive increase in Cronbach's α. Finally, we calculated item-total correlations to evaluate the extent to which each item predicts the overall scale score. Items were removed from the factor structure if the item-total correlation did not fall between .70-.90 (moderate correlations), which resulted from strong internal consistency (Portney & Watkins, 2009). In fact, high item-total correlations ($> .90$) may be indicative of a redundant item and low item-total correlations ($< .70$) may be indicative of a different trait. Therefore the item-total correlations criterion was set as .70-.90 (Portney & Watkins, 2009).

Correlational analysis. Once the item analysis was completed, we evaluated the extent to which the short form of the FTDS predicted the results of the FTDS, through the use of a linear bivariate correlation coefficient (Pearson's r or Spearman's rho, as appropriate). The advantage of a correlational analysis is that it provides the direction and degree to which two variables are related. Nonetheless, Pearson's *r* is criticized for generally producing overestimated measures of the relationship between variables (Kim, 2013). An alternative method for comparing measures is for example the Bland-Altman plot. Bland-Altman plots describe the agreement between two quantitative measures, providing the degree to which the measures are different from one another. Bland-Altman plots require that the limits of expected agreement be defined a-priori based on the clinical goals of the researcher (Kim, 2013). However, despite its limitations, and because of the exploratory nature of the study, the student researcher conducted a correlational analysis. (For a detailed description of the methods please refer to Appendix D).

2.4 Results

Descriptive Statistics

Table 2.1 displays the driver and caregiver demographics. On average, the drivers were older (*M* = 73 years, *SD* = 5.35, range 65-85 years) than their caregivers (*M* = 62 years, *SD* = 14.76, range 18-85 years), and more than half of the drivers were male. The majority of the drivers and their caregivers were white. More than 70% of the older drivers and their caregivers had post-secondary education and at least half of both the older drivers and their caregivers lived with a spouse or partner. More than 80% of the older drivers passed the on-road assessment.

Table 2.1

Demographics of Older drivers and their Caregivers

	Older drivers		Caregiver	
	Frequency (*f*)	Percentage (%)	Frequency (*f*)	Percentage (%)
Gender				
Male	110	55.0	55	27.5
Female	90	45.0	145	72.5
Race				
White	177	88.5	180	90.0
African American	12	6.0	12	6.0
Other	11	5.5	8	4.0
Education				
College/University	114	57.0	93	46.5
Vocational/Associate degree	43	21.5	75	37.5
≤ High school	43	21.5	32	16.0
Living with a partner/spouse	129	64.5	111	55.5
On-road Assessment				
Pass	169	84.5		
Fail	31	15.5		

Exploratory Factor Analysis

Bartlett's test of sphericity was statistically significant, and the KMO coefficient was 0.92, suggesting that the dataset was factorable. Evaluation of the scree plot and the results of the parallel analysis suggested that a three-factor model was the most interpretable factor model for the data. Congruent with the philosophical underpinnings of the fitness to drive construct, this three-factor model represented the person, environment, and vehicle domains. The first factor contained 23 items and was labeled *Complex Driving Tasks*. The second factor contained nine items and was labeled *Routine Driving Tasks;* and the third factor was labeled *Visual Scanning* and contained six items. Table 2.2 presents the 38 items and their factor loadings for the three-factor model.

Table 2.2

Factor loadings based on principal components analysis for the three-factor model

Item	Factor 1: Complex Driving tasks	Factor 2: Routine Driving Tasks	Factor 3: Visual Scanning
51. Drive in a thunderstorm with heavy rains and wind?	.84		
48. Drive at night on a dark road with faded or absent lane lines?	.84		
50. Turn left across multiple lanes when there is no traffic lights	.83		
34. Pass (overtake) a larger vehicle such as a RV?	.80		
25. Make a left hand turn crossing multiple lanes and entering traffic (with no lights or stop signs)?	.80		
33. Pass (overtake) a car in the absence of a passing lane	.80		
47. Drive when there is fog?	.75		
49. Drive when there is glare or the sun is in his or her eyes?	.70		
52. Control his or her car on a wet road?	.68		
45. Drive at night?	.68		
44. Drive in an unfamiliar area?	.66		
32. Drive in dense traffic (such as rush hour)?	.66		
22. Drive with surrounding tractor-trailers (transport trucks)?	.64		
39. Control the car to avoid collisions?	.64		
24. Use a map while driving?	.64		
23. Merge onto a highway?	.63		
42. Drive when upset (anxious, worried, sad or angry)?	.61		
46. Avoid dangerous situations (such as car opening, car pulling out etc.)?	.60		
43. Stay focused on driving when there are distractions?	.55		

10. Drive on a highway with two or more lanes in each direction?	.51		
13. Change lanes in moderate traffic?	.47		
41. Alter his or her driving in response to changes in health?	.43		
11. Keep up with the flow of traffic?	.40		
18. Enter the flow of traffic when turning right?		.63	
16. Maintain lane when turning (not cut corners or go wide)?		.61	
1. Stay in the proper lane?		.59	
27. Stay within the lane markings unless making a lane change?		.53	
19. Share the road with vulnerable road users such as bicyclists?		.47	
28. Stay within the proper lane in the absence of road features?		.45	
6. Obey varied forms of traffic lights (such as green arrow for turn lane or flashing lights)		.43	
15. Brake at a stop sign so car stops completely before the marked line		.41	
9. Drive in light rain?		.40	
4. Check car mirrors when changing lanes			.70
29. Keep distance between his or her car and others?			.65
2. Check for a clear path when backing out from a driveway/ parking space?			.62
30. Look left and right before crossing an intersection?			.57
3. Use the car controls (such as the turn signals)?			.55
21. Check blind spots before changing lines			.55

Table 2.3 displays the total variance and items removed from the factors for the final model. These items were removed because they did not meet the factor loading criteria or they loaded onto two or more factors.

Table 2.3

Extracted factor model's total variance and removed items

	Total Variance	Items with low factor loadings (< 0.40)	Items loading onto two or more factors
Three-factor model	56.2%	5, 7, 8, 14, 17, 31	12

Item Analysis

Scale reliability was high for factors 1 (α = .96, 95% CI [.95, .97]), 2 (α = .88, 95% CI [.85, .90]), and 3 (α = .88, 95% CI [.85, .90]). For each factor, the alpha-if-deleted values were lower than each factor's corresponding Cronbach α, resulting in no removal of items. For item-total correlations, 16 items from factors 2 and 3 (representing "lower" levels of driving ability) were below the .70 criterion. As such, these items were removed, resulting in a short form consisting of 22 items. This short form was entitled FTDS short form version 1 (FTDS V1). The FTDS V1 was made up of 19 out of 22 items in factor 1, 2 out of 9 items in factor 2, and 1 out of 6 items in factor 3. The FTDS V1 had an underrepresentation of items from factors 2 and 3 and potentially as a result of the larger number of items included from factor 1, leading to lower item-total correlations for factors 2 and 3. To include more items from factors 2 and 3, the item-total correlation criterion was decreased to ≥ .60. The theoretical rational behind increasing the representation of items from factors 2 and 3 is discussed below.

The primary study by Classen et al. (2015) presented a logit scale illustrating FTDS item difficulties. Close examination of the items on the scale indicated that all of the factor 1 items were towards the top end of the logit scale representing the most difficult items. Factor 2 contained items in the middle of the logit scale, representing intermediate difficulty items. Factor 3 contained items at the very bottom of the scale, representing the easiest items in the FTDS. Therefore, the items in factor 1 *Complex Driving Tasks* require the highest level of ability to complete the driving task, whereas factors 2 *Routine Driving Tasks* and 3 *Visual Scanning* require intermediate and lower levels of driving ability, respectively. To complete tasks requiring higher levels of ability like those in factor 1, the driver must be able to complete lower level tasks first, as presented by Michon's model of driving behaviours, and used in the initial development of the FTDS (Michon, 1984; Classen et al., 2010).

Michon's model hierarchically organizes driving into three levels of abilities known as operational, tactical and strategic. As each level increases, the ability needed to complete the driving task increases as well (Barkley, 2004). Michon's model can be applied to the three factors of the selected model. The operational level, the first and lowest level of the hierarchy, involves using fundamental functions such as attention, similar to the items in factor 3 *Visual Scanning*. The second tier, tactical, deals with routine choices and decisions drivers make throughout a drive, similar to items in factor 2, *Routine Driving Tasks*. Last but not least, the highest level of the hierarchy known as the strategic level involves drivers planning the drive strategically before and during it, similar to the items in factor 1, *Complex Driving Tasks*. Michon's model uses the premise that to progress in the hierarchy, drivers must be capable of completing lower level tasks first. Therefore, the under-representation of items demonstrating the lower level tasks in FTDS V1 may introduce bias, where the items are only applicable to high ability drivers. Furthermore, a complete range of driver abilities would not be represented in the FTDS V1, and lastly, the three factors would not be adequately represented. Therefore, the criterion was extended to include item-total correlations that were ≥ .60. As a result of this extension, a second short form, entitled FTDS short form version 2 (FTDS V2), comprising of 32 items was constructed. FTDS V2 had two more items in factor 1, five more items in factor 2, and three more items in factor 3. As such, FTDS V2 includes more items from factors 2 (5 items) and 3 (3 items) than V1. Table 2.4 reports the item-total correlations of factors 1-3 for FTDS V1 and V2.

Table 2.4

Item-total correlations for FTDS V1 and V2

	Item-total Correlations for FTDS V1			Item-total Correlations for FTDS V2		
Item	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
52. Control his or her car on a wet road?	.80			.80		
50. Turn left across multiple lanes when there is no traffic lights	.79			.79		
33. Pass (overtake) a car in the absence of a passing lane	.78			.78		
39. Control the car to avoid collisions?	.78			.78		
32. Drive in dense traffic (such as rush hour)?	.78			.78		
46. Avoid dangerous situations (such as car opening, car pulling out etc.)?	.78			.78		

48. Drive at night on a dark road with faded or absent lane lines?	.78		.78	
47. Drive when there is fog?	.76		.76	
34. Pass (overtake) a larger vehicle such as a RV?	.75		.75	
43. Stay focused on driving when there are distractions?	.74		.74	
51. Drive in a thunderstorm with heavy rains and wind?	.74		.74	
44. Drive in an unfamiliar area?	.74		.74	
23. Merge onto a highway?	.72		.72	
13. Change lanes in moderate traffic?	.72		.72	
22. Drive with surrounding tractor trailers (transport trucks)?	.72		.72	
45. Drive at night?	.72		.72	
25. Make a left hand turn crossing multiple lanes and entering traffic (with no lights or stop signs)?	.71		.71	
42. Drive when upset (anxious, worried, sad or angry)?	.70		.70	
10. Drive on a highway with two or more lanes in each direction?	.70		.70	
49. Drive when there is glare or the sun is in his or her eyes?			**.66**	
41. Alter his or her driving in response to changes in health?			**.65**	
16. Maintain lane when turning (not cut corners or go wide)?		.78		.78
9. Drive in light rain?		.70		.70
27. Stay within the lane markings unless making a lane change?				**.69**
18. Enter the flow of traffic when turning right?				**.66**

19. Share the road with vulnerable road users such as bicyclists?		**.64**	
1. Stay in the proper lane?		**.64**	
28. Stay within the proper lane in the absence of road features?		**.64**	
4. Check car mirrors when changing lanes	.81		.81
30. Look left and right before crossing an intersection?			**.69**
21. Check blind spots before changing lines			**.65**
2. Check for a clear path when backing out from a driveway/ parking space?			**.62**

Note. Bolded item-total correlations highlight the items that were added to the short form after the extension, which included values between .60-.69.

Correlational Analysis

The non-significant Shapiro-Wilk tests for the FTDS V1 (W = 70.00, p >.05) and V2 (W = 70.00, p >.05) indicated that the sample had a normal distribution. Accordingly, the concurrent validity of the short form was evaluated against the FTDS using Pearson's r. FTDS V1 (r= .99) had an excellent relationship and FTDS V2 (r = 1.00) had a perfect relationship.

2.5 Discussion

This study used EFA and item analysis to develop a short form FTDS, and employed correlational analysis to quantify the concurrent validity of the constructed short form against the original FTDS. Consistent with Classen et al. (2010), the present study suggested that a three-factor model best represented the empirical data. Scale reliability results (i.e., Cronbach's alpha) indicated high internal consistency of each factor structure within the three-factor model, and a subsequent calculation of the alpha-if-deleted for each item suggested that further removal of items would not reduce Cronbach's α, or the overall homogeneity of factors.

Item reliability results led to the construction of two short forms, FTDS V1 and V2. For FTDS V1, items that did not meet the criterion (≥.70) were removed because they reflected a weak correlation between the item and overall measure (Portney & Watkins, 2009). However, the removal of items with lower item-total correlations resulted in a short form with inadequate representation of factors 2 and 3 of the chosen factor model. Specifically, the lack of representation of items demonstrating intermediate and lower level tasks in the FTDS V1 may have resulted in a short form that did not encompass all aspects of the fitness to drive construct. Therefore the cut point was extended to include item-total correlations that were ≥ .60. Although, the items may suggest a weaker correlation with the overall measure, the introduction of additional items in the FTDS V2: (1) mitigates the potential bias seen in V1 and (2) introduces more items requiring intermediate and lower level driving abilities, including all aspects of the fitness to drive construct.

Correlational Analysis

Correlational analysis results yielded excellent relationships between the original FTDS and short forms V1 (r = .99, p = .05) and V2 (r = 1.00, p = .05). Like the original FTDS, the short forms are measuring the construct of fitness to drive. The endeavor to construct a short form FTDS was taken so that the time it takes to complete the FTDS may be reduced. Therefore, FTDS V1 would be the better choice of the two versions for the following reasons: (1) displayed excellent concurrent validity, (2) contained only 22 items (10 items less than V2), and (3) will hypothetically take less time to complete than

FTDS V2. However, depending on the user goals, either one of the versions may be useful.

FTDS V1 is comprised of 22 items and may be useful for users who only want to identify older drivers with high levels of driving competency. Since this version mostly includes higher-level (challenging) items that require high levels of competency, it may not be able to differentiate routine drivers who have lower levels of driving competency. However, this assertion needs to be tested empirically. The greater variety of items requiring different levels of driving competency in short form V2 (comprising of 32 items) may allow for differentiation between older drivers, with low, intermediate or high levels of driving competency, but this assertion will also require empirical validation. FTDS V1 and V2 are both substantially shorter than the original 54-item FTDS. Logically, the reduced number of items should also reduce the time it takes to complete the tool.

Limitations

EFA sample size recommendations suggest at least 10 participants for each item in the measure, or 300 participants at the very minimum (Yong & Pearce, 2013). The present study (n = 190) was, therefore, somewhat underpowered for this analysis. This limitation was, however, somewhat mitigated through the use of additional metrics of item and measure quality (i.e., the classical psychometric indices). Furthermore, all factor analytic results were evaluated subjectively by content experts (SC) to ensure that adequate coverage of the content domain was seen in each of the proposed latent factors.

Implications

The short form FTDS may increase the completion rates, usage and impact of the tool. That been said, more development, evaluation and refinement is needed before this short form may be used to supplant the original FTDS. The next step in the development of the short forms will be to demonstrate whether these versions can discriminate between at-risk and not at-risk drivers by predicting on-road assessment outcomes. Predictive validity is necessary for any screening measure to gain acceptance by the general public and clinicians (Cordazzo, Scialfa, & Ross, 2016).

A short form FTDS does not change the tool’s intended role for clinicians. Clinicians such as occupational therapists may be able to use caregiver results of the short form FTDS as an entry point for making fitness to drive decisions. However, a short form will expedite this process. Furthermore, the short form ensures that the most coherent aspects of fitness to drive are addressed, providing clinicians with greater latitude in developing next steps (Widaman et al., 2008).

Lastly, the short form FTDS has the potential to become integrated into health policies regarding older drivers in the form of opportunistic screening available at doctors’ offices and health clinics. Opportunistic screening occurs when a person who meets the criteria of a screening measure is offered the test, regardless of the reason for the visit (Institute for Quality and Efficiency in Health Care, 2013). Once the short form FTDS shows to be a valid tool for predicting on road outcomes, it may be implemented as a helpful tool for identifying at-risk older drivers.

2.6 Conclusion

As the populations continue to age fitness to drive screening will become more relevant. According to the World Health Organization, a population-wide screening measure is beneficial when detecting the condition (e.g., fitness to drive) leads to improved health outcomes and effective intervention is available when the condition is detected early (Institute for Quality and Efficiency in Health Care, 2013). The fitness to drive of older adults is becoming a growing concern. Early detection through the FTDS may help facilitate adequate decisions about continued driving, referral and rehabilitation or driving cessation for older drivers.

2.7 References

Alecxih, L. M. B., Zeruld S., & Olearczyl, B. (2001). *Characteristics of caregivers based on the survey of income and program participation. National Family Caregiver Support Program*. Falls Church, VA: The Lewin Group.

Barkley, R. A. (2004). Driving impairments in teens and adults with attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America*, *27*(2), 233-260.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Brouwer, W. H., & Ponds, R. W. H. M. (1994). Driving competence in older persons. *Disability and Rehabilitation*, *16*(3), 149-161.

Classen, S., Medhizadah, S., & Alvarez, L. (2016). The Fitness-to-Drive Screening Measure©: Patterns and trends for Canadian users. *The Open Journal of Occupational Therapy*, *4(*4*)*, 1-15. doi.10.15453/2168-6408.1227

Classen, S., Velozo, C. A., Winter, S. M., Bédard, M., & Wang, Y. (2015). Psychometrics of the Fitness-to-Drive Screening Measure©. *OTJR: Occupation, Participation and Health*, *35*(1), 42-52. doi.10.1177/1539449214561761

Classen, S., Winter, S. M., Velozo, C. A., Bédard, M., Lanford, D. N., Brumback, B., & Lutz, B. J. (2010). Item development and validity testing for a self- and proxy report: The Safe Driving Behavior Measure. *American Journal of Occupational Therapy*, *64*, 296-305.

Cordazzo, S. T. D., Scialfa, C. T., & Ross, R. J. (2016). Modernization of the driver behavior questionnaire. *Accident Analysis and Prevention*, *87*, 83-91. doi.10.1016/j.aap.2015.11.016.

Eby, D. W., Molnar, L. J., Shope, J. T., Vivoda, J. M., & Fordyce, T. A. (2003). Improving older driver knowledge and self-awareness through self-assessment: The Driving Decisions Workbook. *Journal of Safety Research*, *34*, 371-381. doi10.1016/j.jsr.2003.09.006

Family Caregiver Alliance. (2010). Fact sheet: Selected caregiver statistics. *Family Caregiver Alliance: National Centre on Caregiving*. Retrieved from https://www.caregiver.org/caregiver/jsp/print_friendly.jsp?nodeid=439

Federal Highway Administration. (2014). Highway finance data collection, *U.S. National Department of National Transportation.* Retrieved from https://www.fhwa.dot.gov/policyinformation/pubs/hf/pl11028/chapter4.cfm

Google Inc. (2014). *Google Analytics*. Retrieved from www.google.com/analytics

Hollander, M. J., Liu, G., & Chappell, N. L. (2009). Who cares and how much: The imputed economic contribution to the Canadian healthcare system of middle-aged and older unpaid caregivers providing care to the elderly. *Healthcare quarterly, 12*(2), 42-49.

Institute for Quality and Efficiency in Health Care. (2013). *Benefits and risk of screening tests*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072602/

Kim, H-Y. (2013). Statistical notes for clinical researchers: Evaluation of measurement error 2: Dahlberg's error, Bland-Altman method, and Kappa coefficient. *Restorative Dentistry and Endodontics, 38*(3), 182-185. doi 10.5395/rde.2013.38.3.182

Marotolli, R., & Richardson, E. D. (1998). Confidence in, and self rating of, driving ability among older drivers. *Accident Analysis and Prevention, 30*(3), 331-336.

McPeek, R., Nichols, A. L., Classen, S., & Breiner, J. (2011). Bias in older adults' driving self-assessments: The role of personality. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(6), 579-590. doi:10.1016/j.trf.2011.06.001

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? *Human behavior and Traffic Safety*, 485-520.

Musselwhite, C. B. A., & Shergold, I. (2013). Examining the process of driving cessation in later life. *European Journal of Ageing, 10*(2), 89-100. doi.10.1007/s10433-012-0252-6

National Alliance for Caregiving & AARP. (2015). *Caregivers in the U.S.* Retrieved from http://www.caregiving.org/caregiving2015/

National Highway Traffic Safety Administration. (2013). Traffic safety: data 2013. *U.S. National Department of National Transportation.* Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812169

Nicoletta, N. (2013). Driving characteristics of young and aging population. *Statistics Canada.* Retrieved from http://www.statcan.gc.ca/pub/53f0007x/4224875-eng.pdf

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical practice: applications to practice*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall Health.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health*, *14*(8), 1101-1108. doi.10.1016/j.val.2011.06.003

Widaman, K. F., Little, T. D., Preacher, K. J., Sawalani, G. M. (2008). Chapter 3: On creating and using short forms of scales in secondary research. *Secondary data analysis: An introduction for psychologists (*pp. 39-61). Washington, DC: American Psychological Association.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

Chapter 3

3 Concurrent Criterion Validity of the Short Form Fitness-to-Drive Screening Measure©

3.1 Background

The Fitness-to-Drive Screening Measure© (FTDS) has valid and reliable psychometric properties for identifying at-risk older drivers (Classen, Velozo, Winter, Bédard, & Wang, 2015). Using Google Analytics reports (Google Inc., 2014) Classen, Medhizadah and Alvarez (2016) have identified several areas of the FTDS in need of improvement. For instance, although users are accessing the FTDS, it is being underutilized, potentially impacting the measure's uptake and impact as a screening tool. A short form FTDS has been developed to overcome factors that may contribute to the measure's underutilization, such as the time it takes to complete the tool. The psychometric properties of the FTDS short form look promising, demonstrating reliability and concurrent validity with the FTDS. Now, the next step in the development of the short form is to empirically establish its concurrent criterion validity against the gold standard, on-road assessment. Therefore, this study quantifies the concurrent criterion validity of the short form FTDS using Receiver Operating Characteristic curve (ROC) analysis.

3.2 Literature Review

Fitness-to-Drive Screening Measure©

The FTDS is a free online screening measure, available at www.fitnesstodrivescreening.com. This measure enables proxy raters (e.g., formal/informal caregivers, family members or friends) who have driven with the driver in the last 3 months to identify at-risk older (≥ 65) drivers. The FTDS takes approximately 20 minutes to complete and consists of three sections. For sections, A and B the proxy rater completes demographic questions about the driver and him or herself. For section C, the proxy rater uses their observations of the older driver to rate the driver's difficulty on 54 driving-related items (Appendix A contains the FTDS questionnaire). The 54 items are rated using a Likert scale: 2 = *very difficult*, 3 = *somewhat difficult*, 4 = *a little difficult,* 5 = *not difficult*. Once completed, the measure provides the rater with a classification of the driver, solely based on item responses from section C as: an *at-risk driver,* who has urgent safety concerns that must be addressed immediately; *routine driver,* who has some safety concerns present with early signs of needing intervention; or *accomplished driver,* who may not have safety concerns present at the moment, but might experience difficulty with some challenging driving situations. Subsequently, the rater is provided with recommendations and resources appropriate for their level of risk (Classen, 2015). For example, *at-risk drivers* will be recommended to stop driving for the time being and arrange for a comprehensive on-road assessment. They will also receive related resources such as the website for driving assessment centers close to or in their location. Lastly, the rater is provided with a key form summarizing the results and highlighting the driver's overall areas of difficulty on the 54 driving-related items. Caregivers can use the key form to inform clinicians about areas of challenge for the driver.

Psychometric testing of the FTDS has indicated that the measure is a valid and reliable screening tool for caregivers to identify at-risk older drivers (Classen et al., 2015). However, despite the established psychometric properties and ability to identify at-risk older drivers, many users quit before completing the online measure and as such, do not get to the resources and recommendations pages (Classen, Medhizadah, & Alvarez, 2016). Classen and colleagues (2016) suggested several changes to the FTDS that may facilitate more users to complete it, such as decreasing the time needed to complete the measure by constructing a short form. Therefore, our research team constructed two FTDS short forms that may decrease completion time and which may potentially increase utilization of the FTDS as per its original intent.

The research team constructed the first FTDS short form by reducing the number of items in section C of the FTDS using exploratory factor analysis and item analysis. This short form consisted of 22 items and was entitled FTDS short form version 1 (FTDS V1). However, due to the lack of representation of items from the different factors in this version, the research team constructed a second short form with 32 items, entitled FTDS short form version 2 (FTDS V2). For further information on the development of the FTDS short forms, please refer to chapter 2 of this thesis. The concurrent validity of FTDS V1 (r = .99, p = .05) and V2 (r = 1.00, p = .05) with the FTDS denoted excellent to perfect relationships. The correlational analysis results indicated that both FTDS short forms may enable caregivers (i.e., spouses, family members, friends, or informal/formal caregivers) to identify at-risk older drivers. Thus, the next step in the development of the FTDS short forms, before caregivers can use the measures, is to assess if the two versions can correctly classify older drivers who had passed or failed the on-road assessment. Therefore, the FTDS V1 and V2 will need to demonstrate concurrent criterion validity with the gold standard, on-road assessment.

ROC curve

Concurrent criterion validity is the most relevant type of validity when evaluating the usefulness of a clinical tool (Kimberlin & Winterstein, 2008). Concurrent criterion validity is assessed when the outcomes of the clinical tool, e.g., FTDS short form, is compared to a reference test, e.g., on-road assessment that provides an accurate representation of the outcome measure. Furthermore, for the FTDS short form to be useful in a clinical setting, it must have a quantifiable score as a criterion for drivers who pass vs. fail the on-road assessment (Shechtman, Awadzi, Classen, Lanford, & Joo, 2010). The relationship between outcomes of the short form and on-road pass/fail outcomes can be examined by using ROC analysis.

ROC analysis emerged in the early 1950's when signal detection theory and radar technology merged (Streiner & Cairney, 2007). ROC curves were used to separate the "true signals" picked up by the radar from the "noise" (Streiner & Cairney, 2007, p.122). Essentially, by increasing the gain of the radar more signals were picked up, but it also increased the noise and the possibility of misinterpreting the noise as a true signal. On the other hand, decreasing the gain of the radar meant only strong signals were picked up (decreased noise). This made it unlikely to falsely label noise as a true signal but may have also lead to many true signals being missed. Therefore, at some point increasing the gain of the radar would prove to be counter-productive, as the noise would outweigh the true signals. Soon after, other disciplines such as psychology recognized the usefulness of ROC curves and adapted the methodology for such work (Hajian-Tilaki, 2013).

In the last 40 years, ROC analysis has become a popular method for evaluating the accuracy of "diagnostic" tests. This is because ROC analysis is not sample dependent, and provides a summary of the indices needed for a complete description of the measure's ability to discriminate between two groups (Hajian-Tilaki, 2013; Krzanowski & Hand, 2009). Therefore, ROC analysis was used in this study.

A ROC curve is generated by plotting the rate of true positives (sensitivity) against the rate of false positives (1-specificity). *Sensitivity* is the screening tool's (short form FTDS's) ability to predict a fail when the older driver actually failed the on-road assessment (rate of true positives) (Streiner & Cairney, 2007). Thus, sensitivity allows the researcher to determine the proportion of drivers that tested positive (failed the on-road assessment) on the FTDS short form and failed the on-road assessment. Sensitivity is calculated as the (number of true positives) ÷ (the number of true positives + false negatives).

Specificity is the screening tool's (the short form FTDS's) ability to predict a pass when the older driver actually passed the on-road assessment (rate of true negatives) (Streiner & Cairney, 2007). Thus, specificity allows the researcher to determine the proportion of drivers that tested negative (passed the on-road assessment) on the short form FTDS and passed the on-road assessment. Specificity is calculated as (the number of true negatives) ÷ (the number of true negatives + false positives).

The *false positive rate,* also known as a *Type I error,* is when the screening measure predicts a fail although the driver actually passed the on-road assessment, and is calculated as 1- specificity. The *false negative rate,* also known as a *Type II error,* is when the screening measure predicts a pass while the older driver actually failed the on-road assessment and is calculated as 1- sensitivity (Portney & Watkins, 2009; Streiner & Cairney, 2007).

Furthermore, *positive predictive value* (PPV) estimates the proportion of older drivers who actually failed the on-road assessment from the total number of older drivers classified as a fail by the screening measure. PPV is calculated as (number of true positives) ÷ (number of true positives + false positives). The *negative predictive value* (NPV) estimates the proportion of older drivers who actually passed the on-road assessment from the total number of older drivers classified as a pass by the screening measure. NPV is calculated as the (number of true negatives) ÷ (the number of true negatives + false negatives). The closer the PPV and NPV are to 1.00, the better because it reflects a higher probability of correctly classifying older drivers into passing/failing categories. In contrast, lower predictive value indicates a lower probability of correctly classifying older drivers as passing/failing the on-road assessment (Krzanowski & Hand, 2009; Portney & Watkins, 2009). A high PPV would indicate that the screening measure provides a strong estimate of older drivers who actually failed the on-road assessment. Similarly, a high NPV would suggest that the screening measure provides a strong estimate of older drivers who actually passed the on-road assessment (Portney & Watkins, 2009). Other diagnostic indicators, discussed but not used in this study, are positive and negative likelihood ratios. Positive likelihood ratios (PLR) indicate how much to increase the probability of an older adult failing the on-road assessment if classified as a fail by the screening measure. PLR is calculated as (sensitivity) ÷ (specificity -1) (Grimes & Schulz, 2005). Negative likelihood ratios (NLR) indicate how much to decrease the probability of an older adult failing the on-road assessment if classified as a pass by the screening measure. NLR is calculated as (1- sensitivity) ÷ (specificity) (Grimes & Schulz, 2005).

Misclassifications are the number of older drivers that may be erroneously classified by the FTDS short form as either passing or failing the on-road assessment, and is calculated as the sum of the false negative and false positive values. *Error* is the rate of false negative and false positives, represented by the minimum distance between the generated ROC curve and upper left corner of the plot. *Error* is calculated as (1-sensitivity) + (1-specificity) (Krzanowski & Hand, 2009).

The area under the curve (AUC) of the ROC represents the screening measure's ability to differentiate between older drivers who passed/failed the on-road assessment (Portney & Watkins, 2009; Streiner & Cairney, 2007). An AUC value of equal or below .50 indicates that the short form is no better than chance at identifying drivers who passed/failed the on-road assessment. An AUC value above .50, $p < .05$ will demonstrate the short form's ability to correctly discriminate between drivers that passed/failed the on-road assessment. For this study, we used an AUC of ≥ .70 as an acceptable index of discriminability for pass/fail outcomes (Streiner & Cairney, 2007; Zhou, Obuchowski, & Obuchowski, 2002).

ROC curve analysis can be used to determine the optimal cut-point or quantifiable score utilized as a criterion for older drivers who passed/failed the on-road assessment. The optimal cut-point is the point where the overall number of misclassifications (false positives + false negatives) is the lowest (Streiner & Cairney, 2007). Visually, the optimal cut-point is the maximum distance between the generated ROC curve and the diagonal line representing an AUC value of .50. However, empirically this cut-point can be calculated with Youden's index (*J*). This index ranges from 0 to 1, with values closer to 1 suggesting that the overall effectiveness of a cutoff point is relatively large. Values closer to 0 suggest limited effectiveness (Schisterman, Perkins, Liu, & Bondell, 2005; Youden, 1950). Youden's index is calculated as (sensitivity + specificity) -1, with the maximum *J* value used as the criterion for selecting the optimal cutoff point (Schisterman et al., 2005; Youden, 1950).

Gaps in the Literature

No measurement tool can be altered without also changing its inherent validity. Therefore, while the FTDS has shown concurrent criterion validity (.73, $p < .05$), the same cannot be assumed for the FTDS short form. Without empirical evidence, it cannot be assumed that the FTDS short form can also correctly discriminate between older drivers who passed or failed the on-road assessment. Furthermore, screening tools that are not empirically validated may potentially harm those undergoing the screening process by impairing user abilities to accurately screen for fitness to drive and make appropriate follow-up decisions to facilitate fitness to drive or driving cessation. By determining the concurrent criterion validity of the short forms, this study will empirically establish whether, like the FTDS, the FTDS V1 and V2 can discriminate between drivers that passed or failed the on-road assessment.

Rationale and Significance

Using valid screening tools to discriminate between older drivers who passed/failed the on-road assessment can provide valuable insight for caregivers and clinicians, alike. A valid tool can aid caregivers and clinicians in the ability to take appropriate actions concerning driving-related decisions for risk reduction and mitigation. The FTDS has empirical support to do so, but its length may inhibit caregivers to complete it adequately. As such, the FTDS V1 and V2 were developed. In their current formats the FTDS V1 and V2 cannot be used to discriminate between drivers who have passed/failed the on-road assessment because criterion validity has not been established. By empirically testing the concurrent criterion validity of FTDS V1 and V2, we may make inferences to its accuracy, as compared to the actual on-road assessment.

Purpose

The research question for this study was: What is the concurrent criterion validity of FTDS V1 and V2 against the gold standard, on road-assessment? The purpose of this study was to (1) determine whether FTDS V1 and V2 can predict the pass/fail outcome of the on-road assessment (2) establish FTDS V1 and V2's optimal cut-point as the quantifiable score criterion and; (3) quantify FTDS V1 and V2's optimal cut-points associated sensitivity, specificity, PPV, NPV, misclassifications and error.

3.3 Methods

The University of Florida's (UF) Institutional Review Board (IRB201401055) authorized researchers (PI Classen) at the University of Western Ontario (UWO), on July 22, 2016, to use de-identified data, from a previous study conducted with healthy older adults and their caregivers (Classen et al., 2015). The UWO's Office of Human Research Ethics certified this study as a board review exemption (FWA00000121), on March 27, 2015. Appendix B and C contains the ethics letters from the University of Florida and University of Western Ontario.

Design

This study determined the concurrent criterion validity of FTD V1 and V2, using the existing de-identified dataset of caregiver responses to the FTDS and healthy older drivers' on-road pass/fail outcomes.

Participants

In the primary study, healthy older drivers (*n* = 200, age = 65-85 years) and their caregivers (*n* = 200, age = 18-65 years) were recruited from communities in North-Central Florida, United States, and Thunder Bay, Ontario, Canada. Participants were recruited through newspaper advertisements, flyers at local facilities, and word of mouth. Older drivers were included in the primary study if they had a valid driver's license, drove at the time of recruitment, and were physically and cognitively able to take part in the FTDS and on-road assessment. Conversely, older drivers were excluded from the primary study if they had been medically advised not to drive, had experienced seizures, or took medication that impaired their central nervous system. Also, caregivers were included in the primary study if they had observed the older driver in the last three months but were excluded if they displayed signs of physical or mental conditions that could impair their ability to complete the FTDS. The on-road assessments were conducted by a licensed driving school instructor or occupational therapist was also a certified driver rehabilitation specialist (OT-CDRS).

Measure

FTDS short forms. As previously described, the FTDS V1 and V2 only comprise of section C items (items determining difficulty for driving behaviors). FTDS V1 is made up of 22 driving-related items and FTDS V2 is made up of the same 22 driving-related items as V1, plus 10 additional items. As both FTDS V1 and V2 are in the beginning stages of development, an algorithm has not yet been developed to calculate final scores. In the primary study, caregivers responded to items comprising the FTDS using a 4-point Likert scale, ranging from 2 (*very difficult*) to 5 (*not difficult*). Based on their responses, an algorithm (grounded in rasch analysis) produced a final FTDS score in logits. Because an algorithm has not yet been developed for FTDS V1 and V2, this study calculated mean scores of the caregiver responses to items using the 4-point Likert scale. For example, for FTDS V1, if a caregiver responded with 5 (*not difficult*) to 10 items and 4 (*a little difficult*) to another 12 items, the mean score would be [(10 x 5) + (12 x 4)] / 22 = 4.45.

On-road assessment. In the primary study, an OT-CDRS (Florida site) or a licensed driving school instructor (Ontario site) conducted the on-road assessment. The Florida site used a standardized road course with demonstrated reliability (intra-class correlation coefficient [*ICC*] = .94, $p < .05$) and strong correlational relationship between the Global Rating Score's (GRS) four levels (score of 3 = pass, 2 = pass with restrictions or recommendations, 1 = fail with remediation, 0 = fail not remediable), and driving performance scores ($r = .84$, $p < .05$) for older drivers (Justiss, Mann, Stav, & Velozo, 2006). The Ontario site used a demerit point system identical to the Ontario Ministry of Transportation's licensing procedure. For both sites, the on-road assessment GRS scores were further dichotomized into pass (GRS score of 2, 3) or fail (GRS score of 0, 1) outcomes. Interrater reliability between the evaluators at the Florida and Ontario site was 100% (Classen et al., 2010). For this study, the dichotomized pass/fail outcomes of the participants' on-road assessments were used for the ROC analysis.

Data Collection and Management

The research team received de-identified secondary data as an SPSS file (Version 20; IBM Corporation, Armonk, NY). The research student first created two password-protected Excel sheets entitled "FTDS V1" and "FTDS V2". From the SPSS file, the research student copied the 22 FTDS items and caregiver responses and pasted them in the excel file for “FTDS VI.” After that, the research student copied the 32 FTDS items and caregiver responses and pasted them in the excel file for “FTDS V2.” Caregiver responses with missing data were not removed but coded as NA.

For the ROC curve analysis the participants’ mean score on the FTDS V1 and V2 was obtained by calculating the mean of caregiver responses (2 = *very difficult*, 3 = *somewhat difficult*, 4 = *a little difficult*, 5 = *not difficult*) to items comprising each short form. Conversely, ROC analysis requires the use of interval data, which a Likert scale is often assumed not to be representative of. However, the many properties of a Likert scale suggest that it may be presented and used as an interval scale. Specifically, it has been argued in the literature that items within a Likert scale when individually examined are independent and autonomous but when observed as a scale, are demonstrative of a continuous construct (i.e., in the case of the FTDS, the scale is representative of difficulty) that is conceptually and logically linked, exhibiting similar properties to that of interval data (Carifio & Perla, 2007). Furthermore, because the items within a Likert-type scale are ordinal, it is assumed that only non-parametric statistical tests can be carried out. Research using parametric statistical tests with Likert scale data has shown empirical evidence that Likert data exhibit linear and equal intervals (Carifio & Perla, 2007; Labovitz, 1967; Traylor, 1983). This study operated on the assumption that the Likert-scale is demonstrative of interval scale qualities and therefore using the means of caregiver responses to the FTDS short forms was appropriate for the ROC analysis.

Hence, caregiver responses to the Likert scale were summed and then divided by the total number of items within the short form (i.e., 22 items for FTDS V1 and 32 items for V2). For caregiver responses with missing data (coded as NA), the sum of caregiver responses was divided by the total number of items within the short form minus the number of items with missing responses. For example, the mean score for a participant missing responses to 3 items in FTDS V1 was averaged based on 19, not 22 items. Therefore, this study accounted for missing responses. Furthermore, the lowest possible mean score was 2 and the highest possible mean score was 5. All the data were stored on a password-protected server network at UWO and was only accessible to the research team.

Procedure

For FTDS V1 and V2, the research student generated a ROC curve to determine predictive validity with on-road outcomes. To create the curve for FTDS V1, the research student plotted the true positive rate (sensitivity) against the false positive rate (1-specificity). Next, the research student computed the AUC for FTDS V1. The AUC criterion used was ($\geq .70$, $p < .05$). In this study, the *p* value was used as the estimate of precision, but confidence intervals (e.g., CI 95%) could have also been used (Portney & Watkins, 2009). If the AUC met the criteria, Youden's index (*J*) was computed to determine the optimal cut-point (FTDS V1 mean score) or largest *J* value calculated in the index (Schisterman et al., 2005; Youden, 1950). An FTDS V1 score less than the optimal cut-point meant the FTDS V1 predicted the older driver had failed the on-road assessment. Likewise, a mean score greater than or equal to the optimal cut-point on the FTDS V1 meant the driver was predicted to pass the on-road assessment. If the AUC did not meet the criterion ($\geq .70$, $p < .05$), no further analysis was carried out. For the optimal cut-point the research student identified the associated sensitivity, specificity, PPV, NPV, misclassifications, and error. The AUC and corresponding ROC indicators were computed for FTDS V2 in the same way as described above for FTDS V1.

Data Analysis

R Statistics software version 3.1.2 (R Core Team, 2015) was used for the ROC curve analyses. First, the research student uploaded the excel file "FTDS V1" on to R. Next, for FTDS V1 the research student used the R script for ROC analysis to run the computation and obtain the curve's sensitivity, specificity, and AUC values. Once it was determined that the AUC for V1 met the a-priori criteria of $\geq .70$, $p \leq .05$, sensitivity and specificity from the ROC curve were inputted into Youden's index. The pair with the largest *J* value was identified, and its accompanying FTDS V1 score was selected as the optimal cut-point. It must be noted that by using the optimal cut-point (with proportionate sensitivity and specificity), there is an implicit assumption that the "cost" (financial, emotional, physical) of making a Type I error is the same as making a Type II error. This is not always the case as the "cost" of an error may have different implications in different setting. As such it is up to the user to decide the ideal combination of sensitivity and specificity. Different cut-points, representing different sensitivity and specificity values can be used to provide different outcomes to match the priorities of the users. Because the FTDS short form is in its early stages of development and not intended to be used in clinics yet, we selected the optimal cut-point, which is the point on the ROC curve where the overall number of misclassifications (false positives + false negatives) is the lowest. Next, the research student calculated the PPV, NPV, misclassifications and error for FTDS V1's optimal cut-point using previously mentioned equations. The exact procedures were repeated for determining the ROC and it's associated values, as described above, for FTDS V2.

3.4 Results

Participant demographics

Two hundred (110 male and 90 female) older drivers and 200 caregivers (55 male and 145 female) participated in the primary study. On average, the caregivers were younger (*M* = 62 years, *SD* = 14.76 years, range = 18-85 years) than the older drivers (*M* = 73 years, *SD* = 5.35 years, range = 65-85 years). From the 200 older drivers, 169 passed (84.5%) the on-road assessment, while 31 drivers (15.5%) failed. More than half of the drivers (64.5%) and caregivers (55.5%) lived with a spouse or partner. The majority of drivers (88.5%) and caregivers (90.0%) were racially White. African American participants made up 6% of the drivers and caregivers. The remaining 5.5% of drivers and 4.0% of caregivers were comprised of other races. Drivers and caregivers with various levels of education participated in the primary study. Specifically, 57.0% of drivers had a college or university degree; 21.5% had a vocational or associate degree; and 21.5% had high school education or less. Forty six percent of caregivers had a college or university degree; 37.5% had a vocational or associate degree; and 16.0% had high school education or less. This data have been reported in chapter 2 (Table 2.1) of this thesis and in the primary study (Classen et al., 2015).

ROC Curve

FTDS V1. Figure 3.1 presents the ROC curve and AUC for FTDS V1. The AUC for the measure was acceptable at a magnitude of .73, $p < .05$. Youden's index results indicated that the optimal cut-point was a mean score of 4.88 (range = 2-5) on the FTDS V1.

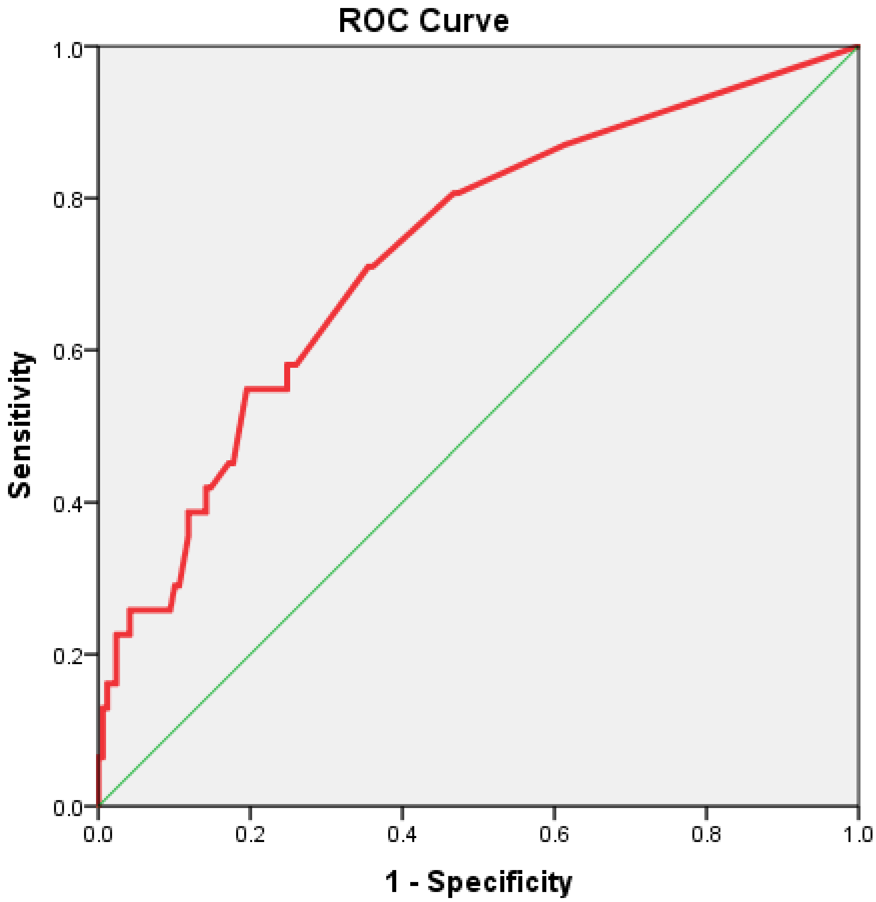

Figure 3.1. ROC Curve for FTDS V1
Note. AUC = .73, $p < .05$

Table 3.1 presents FTDS V1's pass/fail classification of older drivers based on the optimal cut-point. The optimal cut-point of 4.88 yielded a sensitivity of .71 (71%). This means that 22 out of the 31 drivers who failed the on-road assessment obtained a mean score of less than 4.88 on the FTDS V1. Additionally, the specificity value was .65 (65%), indicating that 110 out of the 169 drivers who passed the on-road assessment had a mean score equal to or greater than 4.88 on the FTDS V1. Based on the optimal cut-point of 4.88, PPV was .27 (27%), and NPV was .92 (92%). The PPV indicated that 22 out of the 81 older drivers who had a mean score less than 4.88 on the FTDS V1 actually failed the on-road assessment. The NPV suggested that 110 of the 119 older drivers who had a mean score equal to, or greater than 4.88 on the FTDS V1, actually passed the on-road assessment. At the optimal cut-point there were 68 (out of 200) misclassifications, meaning more than 30% of the older drivers were incorrectly classified as passing (when they actually failed) or failing (when they actually passed) the on-road assessment. The error rate, that is the rate of misclassifications when sensitivity and specificity are given equal weight, was .64 (64%).

Table 3.1

FTDS V1's classification of older drivers based on the optimal cut-point of 4.88

	On-road assessment outcomes		
FTDS V1 outcomes	Fail	Pass	Total
Fail	22	59	81
Pass	9	110	119
Total	31	169	200

Note: FTDS V1 = FTDS short form version 1. Sensitivity = .73, Specificity = .65, PPV = .27, NPV = .92, Misclassifications = 68, Error = .64

FTDS V2. Figure 3.2 presents the ROC curve and AUC for FTDS V2. The AUC for the measure was acceptable at a value of .75, $p < .05$. Youden's index results indicated that the optimal cut-point was a mean score of 4.88 (range = 2-5) on the FTDS V2.

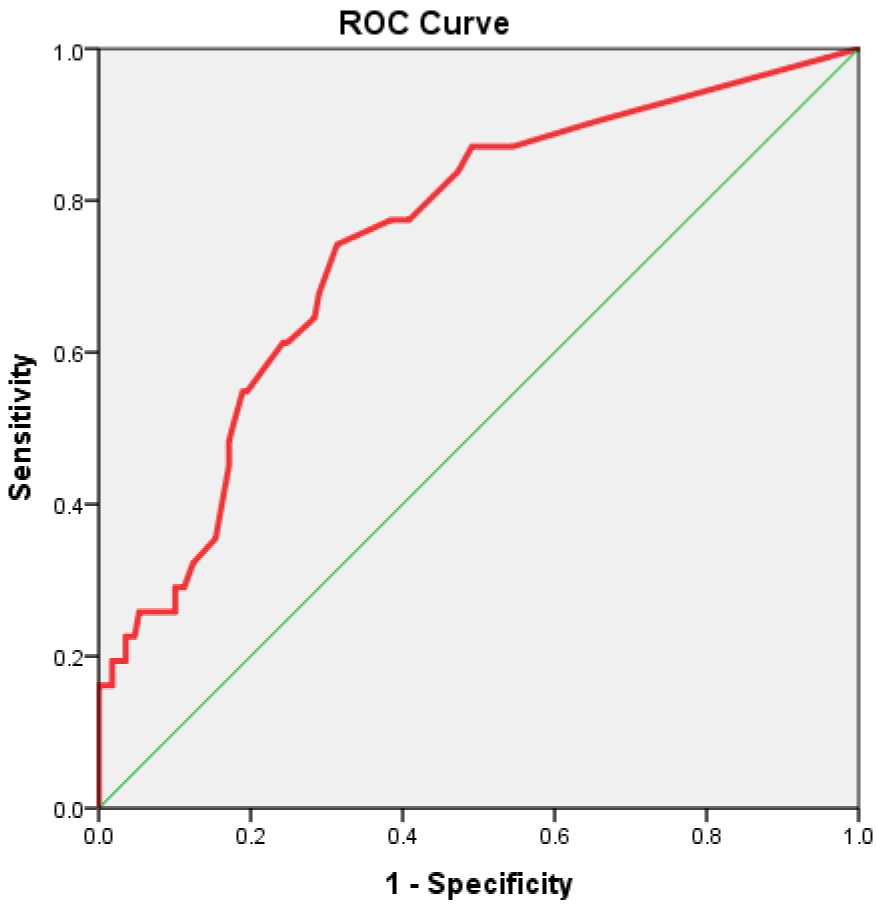

Figure 3.2. ROC Curve for FTDS V2
Note. AUC = .75, $p < .05$

Table 3.2 displays FTDS V2's pass/fail classifications of older drivers based on the optimal cut-point of 4.88. This optimal cut-point yielded a sensitivity of .74 (74%) and specificity of .69 (69%). The sensitivity indicated that 23 out of the 31 drivers who failed the on-road assessment had a mean score of less than 4.88 on the FTDS V2. The specificity indicated that 116 out of 169 drivers who passed the on-road assessment had a mean score greater than or equal to 4.88 on the FTDS V2. Based on this optimal cut-point, the PPV was .30 (30%) and the NPV was .93 (93%). The PPV indicated that 23 out of the 76 drivers who had a mean FTDS V2 score of less than 4.88 actually failed the on-road assessment. The NPV indicated that 116 of the 124 older drivers who had a FTDS

V2 mean score of greater or equal to 4.88 actually passed the on-road assessment. At the cut-point of 4.88,-61 (out of 200) drivers were misclassified as passing (when they actually failed) or failing (when they actually passed) the on-road assessment. Overall, the FTDS V2 had an error rate of .57 (57%) when sensitivity and specificity were given equal weight.

Table 3.2

FTDS V2's classification of older drivers based on the optimal cut-point of 4.88

	On-road assessment outcomes		
FTDS V2 outcomes	Fail	Pass	Total
Fail	23	53	76
Pass	8	116	124
Total	31	169	200

Note: FTDS V2 = FTDS short form version 2. Sensitivity = .74, Specificity = .69, PPV = .30, NPV = .93, Misclassifications = 61, Error = .57

3.5 Discussion

Using ROC analysis, this study (1) determined whether FTDS V1 and V2 predicted pass/fail outcomes of an on-road assessment; (2) established the optimal cut-points for FTDS V1 and V2; and (3) quantified both version's accompanying sensitivity, specificity, PPV, NPV, misclassifications and error rate at the optimal cut-point.

Participant Demographics

In this study, the caregiver characteristics were representative of Canadian and American caregiver populations. Similar to caregiver trends in Canada and the U.S., the majority (72.5%) of caregivers in this study were female and younger (*M* = 62 years) than the older drivers (*M* = 73 years) they cared for (Family Caregiver Alliance, 2010;

Hollander, Liu, & Chappel, 2009). Furthermore, the average age of caregivers (62 years) in the study was comparable to the literature (63 years) (Family Caregiver Alliance, 2010).

ROC Curve

FTDS V1. For the FTDS V1, the AUC (73%) indicated that the measure is overall, an acceptable predictor of pass/fail on-road outcomes. The optimal cut-point of 4.88 was the criterion score used to classify older drivers as either passing or failing the on-road assessment. Based on the optimal cut-point the FTDS V1 had higher sensitivity (71%) than specificity (65%). Thus, the measure had a higher probability of correctly classifying older drivers who actually failed the on-road assessment as failing, than classifying older drivers who actually passed as passing. Specifically, the measure is more likely to correctly identify those who have a mean score of less than 4.88 to fail than those who have a score equal to or greater than 4.88 to pass. Furthermore, lower specificity at the optimal cut-point suggests there are more false positive (Type I) than false negative (Type II) errors. That is, out of the 68 misclassifications, 87% (59/68) were incorrectly classified as failing when they actually passed the on-road assessment or had a score of less than 4.88 and passed on-road assessment. Some of the consequences of type I error may include stress, anxiety, and financial hardships due to potentially being unfit to drive and the recommended next steps. For example, completing a comprehensive driving evaluation can be time-consuming (e.g., 3 hours to complete) and costly (e.g., $900), often with no third party reimbursement, potentially resulting in unnecessary stress and financial hardship for the driver (Weaver & Bédard, 2012). Only 7% (9/68 of the misclassifications were Type II errors. That is, drivers were misclassified by the measure as passing the on-road assessment when they actually failed. Type II errors can also have negative consequences for the older driver. For instance, older drivers may continue to drive although they are not fit to do so, increasing the risk of getting into a crash and injuring or killing the driver or other road users (Weaver & Bédard, 2012). The relative cost of Type II errors (a life) compared to Type I errors (money) is much higher.

The NPV (.92) was higher than the PPV (.27), indicating that 92% of the pass classifications made by the FTDS V1 were correct, whereas only 27% of the fail classifications made by the FTDS V1 were correct. The FTDS V1 had high sensitivity but low PPV. This pattern may result from the imbalance of pass (169) and fail (31) outcomes in the sample. When the dichotomous outcomes are very different in number, as in this case, most standard algorithms favor the larger group (pass outcomes) resulting in poorer accuracy in the minority group's (fail outcomes) predictive value. As a result spectrum bias can be introduced (Lin & Chen, 2012; Weaver & Bédard, 2012).

At the optimal cut-point, the FTDS V1 has an error rate of 65%, indicating the rate of misclassifications when sensitivity and specificity are given equal weight. The overall results of the ROC analysis indicated the FTDS V1 has demonstrated concurrent criterion validity for predicting older driver pass/fail outcomes in this sample, at one point in time, for a limited geographic region. The student researcher must also heed to caution, especially with the future or further development of the short form as error and misclassifications of drivers are apparent.

FTDS V2. For the FTDS V2, the AUC (75%) indicated that the measure is an adequate predictor of pass/fail on-road outcomes. Interestingly, for the FTDS V2, the optimal cut-point was also 4.88. Based on the optimal cut-point the FTDS V2 had higher sensitivity (75%) than specificity (69%). Similar to the FTDS V1, V2 had a higher probability of correctly classifying older drivers who actually failed the on-road assessment as failing, than correctly classifying older drivers who actually passed the on-road assessment as passing. Therefore, FTDS V2 is more likely to correctly identify those who have a mean score of less than 4.88 to fail than those who have a mean score equal to or greater than 4.88 to pass. The lower specificity at the optimal cut-point also signifies that there are more false positive (Type I) than false negative (Type II) errors. This means that more drivers that passed the on-road assessment were erroneously predicted to fail than drivers that had failed the on-road assessment and were erroneously predicted to pass. FTDS V2 had 61misclassifications, eighty-seven percent (53/61) of the misclassifications were Type I errors (false positive) and 13% (8/61) of the misclassifications were Type II errors (false negatives). Similar to FTDS V1, Type I (e.g., stress, anxiety, financial hardship) and Type II (crash risk and killing or injuring the driver or other road users) errors can have negative consequences for the driver, with the cost of Type II errors being relatively higher than Type I errors.

Akin to FTDS V1, FTDS V2 also had a high NPV (.93) and low PPV (.30). Based on the FTDS V2, ninety-three percent of the pass classifications made were correct, while only 30% of the fail classifications were correct. The lower PPV for the FTDS V2 may have been the result of spectrum bias, introduced by the imbalance between the numbers of pass (169) and fail (31) outcomes in the sample (Lin & Chen, 2012; Weaver & Bédard, 2012). The FTDS V2 has an error rate of 59% at the optimal cut-point, indicating the rate of misclassifications when sensitivity and specificity are given equal weight. The FTDS

V2 has demonstrated concurrent criterion validity for predicting older driver pass/fail outcomes in this sample. The validity results are limited to the geographic region and point in time of the sample used for the study. Furthermore, there are errors and misclassifications present and the student researcher must proceed with caution in the future or when further developing the short form.

Both the FTDS V1 and V2 demonstrated good concurrent criterion validity with the gold standard on-road assessment, but the FTDS V2 is a more accurate predictor of older driver's fitness to drive than V1. Specifically, the FTDS V2 correctly classified older drivers as a pass when they actually passed or fail when they actually failed the on-road assessment than V1. Furthermore, FTDS V2 misclassified fewer older drivers than FTDS V1. Nonetheless, any misclassification or error when predicting pass/fail outcomes of older drivers can negatively impact older drivers and their loved ones. Despite the presence of error, this study's results suggest that perhaps by using the FTDS V2 in conjunction with other available clinical information, may provide plausible fitness to drive decisions. For instance using the FTDS V2 alongside sound clinical reasoning, a thorough history of the client and/or the client/caregiver goals may result in appropriate fitness to drive decisions.

Clinical implications and next steps

The FTDS V1 and V2 are clinically meaningful measures that can discriminate between drivers that may pass/fail an on-road assessment. The initial validity testing of the measures indicate that the FTDS short forms have the potential to be incorporated and used in clinical settings in the future. In a clinical setting, the FTDS short form (completed by caregivers) alongside other clinical information (e.g., client history/clinical reasoning) can help healthcare professionals screen for at-risk drivers and make informed initial fitness to drive decisions. However, before the short forms can be used or even incorporated into clinical settings: (1) the scoring algorithm must be developed; and (2) the measures with the new algorithm must be qualitatively and quantitatively tested among caregivers and older drivers.

Strengths and Limitations

The FTDS short forms have been developed, and initial validity testing has been carried out. However, there are many aspects of the short forms that still need to be developed or psychometrics properties that still remain unknown and need to be empirically validated. Currently, no scoring algorithm has been developed for the FTDS short forms. Although the preliminary findings show that the FTDS short forms are valid, validity testing will need to be redone with the new algorithm in place. Furthermore, this study only looked at secondary data from UF, and so it is still unknown whether the short forms can actually predict on-road outcomes in different settings and with different evaluators. The FTDS short forms are also not web-based, reducing the accessibility of the measure for users. The short forms were initially developed to decrease the time it takes to complete the measure, but it has yet to be empirically validated. Lastly, it has been suggested that the FTDS short forms may be valuable measures for screening for at-risk older drivers, however, how caregivers and drivers or health care professionals will respond to the measures still needs to be explored.

The use of a ROC curve analysis for determining the concurrent criterion validity of the FTDS short forms was advantageous. It provided a summary of the information needed for a comprehensive description of both measures’ ability to discriminate pass/fail outcomes. Unlike other methods of calculating accuracy/predictive validity (e.g., percentage of correct pass/fail outcomes in the entire sample), ROC analysis is not sample dependent (Krzanowski & Hand, 2009). This study provided preliminary information about the psychometric properties of the short forms that can be used as the foundation for continued development of the measures.

3.6 Conclusion

Concurrent criterion testing of the FTDS V1 and V2 against the gold standard on-road assessment indicates that the FTDS V1 and V2 can validly identify drivers who pass/fail an on-road assessment. The potential that these short forms may have clinical applicability exists. Based on this study, the FTDS V2 may be the better choice as it has better sensitivity, specificity, PPV, NPV values, and lower misclassifications and error than V1. The next step in the development of the measures is to develop a scoring algorithm and qualitatively and quantitatively test the short forms among caregivers and older drivers.

3.7 References

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert response formats and their antidotes. *Journal of Social Sciences, 3*(3), 106-116. doi.10.3844/jssp.2007.106.116

Classen, S. (2015). Fitness-to-Drive Screening Measure© User Manual. *University of Florida*. Retrieved from http://www.fitnesstodrivescreening.com/ca/

Classen, S., Medhizadah, S., & Alvarez, L. (2016). The Fitness-to-Drive Screening Measure©: Patterns and trends for Canadian users. *The Open Journal of Occupational Therapy, 4(*4*),* 1-15. doi.10.15453/2168-6408.1227

Classen, S., Velozo, C. A., Winter, S. M., Bédard, M., & Wang, Y. (2015). Psychometrics of the Fitness-to-Drive Screening Measure©. *OTJR: Occupation, Participation, and Health*, *35*(1), 42-52. doi.10.1177/1539449214561761

Classen, S., Winter, S. M., Velozo, C. A., Bédard, M., Lanford, D. N., Lutz, B. J. (2010). Item development and validity testing for a self- and proxy report: The Safe Driving Behavior Measure. *American Journal of Occupational Therapy*, *64*, 296–305. doi.10.5014/ajot.64.2.296.

Family Caregiver Alliance. (2010). Fact Sheet: Selected caregiver statistics. *Family Caregiver Alliance: National Centre on Caregiving*. Retrieved from https://www.caregiver.org/caregiver/jsp/print_friendly.jsp?nodeid=439

Google Inc. (2014). *Google Analytics*. Retrieved from www.google.com/analytics

Grimes, D. A., Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet*, *365*(9469), 1500-1505. doi. 10.1016/S0140-6736(05)66422-7

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627-635.

Hollander, M. J., Liu, G., & Chappell, N. L. (2009). Who cares and how much: The imputed economic contribution to the Canadian healthcare system of middle-aged and older unpaid caregivers providing care to the elderly. *Healthcare Quarterly*, *12*(2), 42-49.

Justiss, M., Mann, W., Stav, W., & Velozo, C. (2006). Development of a behind-the-wheel driving performance assessment for older adults. *Topics in Geriatric Rehabilitation*, *22*, 121-128.

Kimberlin, C. J., & Winterstein, A. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*, 2276-2284. doi:10.2146/ajhp070364

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. Broken Sound Parkway, N.W.: Taylor & Francis Group, LLC.

Labovitz, S. (1967). Some observations on measurement and statistics. *Social Forces*, *46,* 151-160. doi:10.2307/2574595.

Lin, W-J., & Chen, J. J. (2012). Class-imbalanced classifiers for high-dimensional data. *Briefing in Bioinformatics*, 1-14. doi: 10.1093/bib/bbs006

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical practice: applications to practice*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall Health.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding youden index to discriminate individuals using blood samples. *Epidemiology*, *16*(1), 73-81. Retrieved from http://www.jstor.org/stable/20486002

Shechtman, O., Awadzi, K. D., Classen, S., Lanford, D. N., & Joo, Y. (2010). Validity and critical driving errors of on-road assessment for older drivers. *American Journal of Occupational Therapy, 64*, 242-251. doi:10.5014/ajot.64.2.242

Streiner, D. L., & Cairney, J. (2007). What's under the roc? An introduction to receiver operating characteristic curves. *The Canadian Journal of Psychiatry, 53*(2), 121-128.

Traylor, M. (1983). Ordinal and interval scaling. *Journal of the Market Research Society, 24*(4), 297-303.

Weaver, B., & Bédard, M. (2012). Assessing the fitness to drive: Practical tips on choosing the right screening tools for your practice. *Canadian Geriatrics Society of CME* 2(*3*), 5-9.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32-35. doi*:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.*

Zhou, X. H., Obuchowski, N. A., & Obuchowski, D. M. (2002). *Statistical methods in diagnostic medicine*. New York, NY: John Wiley and Sons.

Chapter 4

4 Conclusion

This study constructed and validated two FTDS short forms. As previously mentioned the endeavour to develop a short form was initially taken on because users were not completing the tool, potentially due to the time needed to complete the measure. Although the FTDS short forms developed and validated in this study have a reduced number of items, they are far from ready. The short forms constructed in this study are the preliminary building blocks needed for the development of a fully functioning FTDS short form. Initial testing of the concurrent criterion validity of the screening tools determined that the measure can predict on-road pass/fail outcomes and thus can be and should be further developed. If the developed short forms were unable to withstand preliminary validity testing, further constructing the measure would be futile. Now that the initial versions of the FTDS short form have been constructed and shown to be predictive of on-road pass/fail outcomes, the next step is to develop an algorithm for the tools grounded in rasch theory and further validate the measures.

Implications for the occupation of driving

For older adults in Canada and the United States, driving is an important occupation and a meaningful activity of daily living (Stav, 2008; Zur & Vrkljan, 2014). Driving enhances an individual's quality of life and wellbeing by improving ease of transportation and enabling access to community participation settings. Furthermore, driving is a means of mobility that enables older drivers and their loved ones to engage in other important and meaningful occupations linked to the quality of life and health, such as leisure and recreation (Stav, 2008). However, age-related declines, such as visual or cognitive deficits, can impact an older adult's fitness to drive, and may eventually lead to driving cessation. Driving cessation can lead to feelings of isolation and decreased sense of autonomy for both older adults and their family members (Dickerson, Meuel, Ridenour & Cooper, 2014). Therefore, the impact that being fit to drive can have on the occupational engagement of older adults underscores the importance of using screening tools and providing support to identify at-risk drivers and to start early and positive conversations about their fitness to drive abilities. When fully developed, the FTDS short form may help provide such support by promoting a preemptive approach to continued driving, driving interventions, or driving transitions, while supporting, drivers and their loved ones in maintaining the occupation of driving. Furthermore, like the FTDS the short forms may be able to equip caregivers with strategies and resources for decision-making related to continued driving, referral to rehabilitation, or driving cessation for the driver. Allowing older drivers to stay on the road safer for longer.

4.1 References

Dickerson, A. E., Meuel, D. B., Ridenour, C. D., & Cooper, K. (2014). Assessment tools predicting fitness to drive in older adults: A systematic review. *The American Journal of Occupational Therapy, 68*(6), 670-680. doi: 10.5014/ajot.2014.011833

Stav, W. B. (2008). Occupational therapy and older drivers: Research, education, and practice. *Gerontology & Geriatrics Education*, *29*(4), 336-350. doi: 10,1080/02701960802497860

Zur, B. M. & Vrkljan, B. H. (2014). Screening at-risk older drivers: A cross-program analysis of Canadian occupational therapy curricula. *Physical & Occupational Therapy in Geriatrics*, *32*(1), 10-24. doi: 10.3109/02703181.2013.865697

Appendix

Appendix A

Fitness-to-Drive Screening Measure©

Available at www.fitnesstodrivescreeening.com

Instructions:

1. Please answer all 54 questions to the best of your ability.

2. From your observations of the driver over the **past three months**, rate the amount of difficulty for each skill. If you have not observed the driver for a skill, use your best judgment to rate the difficulty the driver would have using one of the following answers:
   - **Very Difficult** - doing it is a major challenge
   - **Somewhat Difficult** - doing it is a moderate challenge
   - **A Little Difficult** - doing it is a minor challenge
   - **Not Difficult** - can do it with ease

3. For each question, please select your answer by clicking on the text or circle.

4. Do not use the back button of the browser, it will not return you to the previous set of questions.

Note the example below:

For the person you are rating, based on the last 3 months, how difficult is it for him or her to...

Start the car?

- ○ Very Difficult
- ○ Somewhat Difficult
- ○ A Little Difficult
- ○ Not Difficult

For the person you are rating, based on the last 3 months, how difficult is it for him or her to...

1. Drive in the proper lane?
2. Check for a clear path when backing out from a driveway or parking space?
3. Use the vehicle controls (such as the turn signals, emergency brake, windshield wipers, or headlights)?
4. Check car mirrors when changing lanes?
5. Read road signs far enough in advance to react (such as make a turn)?
6. Obey varied forms of traffic signals (such as green arrow for turn lane or flashing lights)?
7. Drive and hold a conversation with one or more passengers?
8. Drive with a passenger who is providing driving directions or assistance?
9. Drive in light rain?
10. Drive on a highway with two or more lanes in each direction?
11. Keep up with the flow of traffic?
12. Keep distance from other vehicles when changing lanes?
13. Change lanes in moderate traffic?
14. Drive cautiously (to avoid collisions) in situations when others are driving erratically (such as speeding, road rage, crossing lane lines or driving distracted)?
15. Brake at a stop sign so car stops completely before the marked line?
16. Maintain lane position when turning (not cut corner or go wide)?
17. Back out of parking spots?

18. Enter the flow of traffic when turning right?

19. Share the road with other road users such as bicyclists, scooter drivers, motorcyclists?

20. Drive on graded (unpaved) road?

21. Check blind spots before changing lanes?

22. Drive with surrounding tractor trailers (transport trucks)?

23. Merge onto a highway?

24. Use a paper map while driving?

25. Make a left hand turn crossing multiple lanes and entering traffic (with no lights or stop signs)?

26. Parallel park?

27. Stay within the lane markings unless changing lanes?

28. Stay within the proper lane on roads without road features such as clearly marked lane lines, reflectors or rumble strips?

29. Keep distance between his or her car and others (allow time to react to hazards)?

30. Look left and right before entering an intersection?

31. Drive in a work or construction zone?

32. Drive in dense traffic (such as rush hour)?

33. Pass (overtake) another car on a road without a passing lane?

34. Pass (overtake) a larger vehicle such as a RV, tractor-trailer (transport truck), or dump truck on a road without a passing lane?

35. Drive in an unfamiliar urban area?

36. Control his or her car when going down a steep hill?

37. Exit an expressway, or inter-state highway from the left-hand lane?

38. Drive in a highly complex situation (such as a large city with high-speed traffic, multiple highway interchanges and several signs)?

39. Control the vehicle (brake hard or swerve) to avoid collisions?

40. Drive a different vehicle (such as another person's car or a rental car)?

41. Alter his or her driving in response to changes in health or condition (such as vision, reaction time, fatigue, thinking, joint stiffness, medications)?

42. Drive when upset (anxious, worried, sad or angry)?

43. Stay focused on driving when there are distractions (such as radio, eating, drinking, pet in the car)?

44. Drive in an unfamiliar area?

45. Drive at night?

46. Avoid dangerous situations (such as car door opening, car pulling out, road debris, or an animal darting in front of car)?

47. Drive when there is fog?

48. Drive at night on a dark road with faded or absent lane lines?

49. Drive when there is glare or the sun is in his or her eyes?

50. Turn left across multiple lanes when there is no traffic signal?

51. Drive in a thunderstorm with heavy rains and wind?

52. Control his or her car on a wet road?

53. Drive on a snow covered road?

54. Drive on an icy road?

Appendix B

UF | Institutional Review Board
UNIVERSITY *of* FLORIDA

Health Center Institutional Review Board
FWA00005790

[redacted]

DATE: 7/22/2016

TO: Sandra Winter

[redacted]

FROM: Peter Iafrate, IRB Chairman, University of Florida
Chair IRB-01

IRB#: **IRB201401055**

TITLE: Outcomes of Driver Evaluation and Fitness-to-Drive screening, comparing characteristics of older adult drivers with and without Parkinson's Disease

Approved as Exempt

You have received IRB approval to conduct the above-listed research project. Approval of this project was granted on 7/22/2016 by IRB-01. This study is approved as exempt because it poses minimal risk and is approved under the following exempt category/categories:

4. This research involves the collection or study of existing data*, documents, records, pathological specimens, or diagnostic specimens. These data sources are either publicly available or collected in a manner which subjects cannot be directly or indirectly identified. (* - "existing data" is as of date submitted to IRB)

Approval Includes, but is not limited to:

Special notes to Investigator (if applicable):

Reviewer Notes: 0 Reviewer Notes

Principal Investigator Responsibilities:

The PI is responsible for the conduct of the study. Please review these responsibilities described at: http://irb.ufl.edu/irb01/researcher-information/researcherresponsibilities.html
Important responsibilities described at the above link include:

- Using currently approved consent form to enroll subjects (if applicable)
- Renewing your study before expiration
- Obtaining approval for revisions before implementation
- Reporting Adverse Events
- Retention of Research Records
- Obtaining approval to conduct research at the VA
- Notifying other parties about this project's approval status

Study Team:

The Foundation for The Gator Nation
An Equal Opportunity Institution

Appendix C

Western Research

Research Ethics

March 27th, 2015

Prof. Sherrilene Classen, PhD
[redacted]

Re: Letter of support for Non-Medical Research Ethics Board Review exemption
Dear Prof. Classen,

Thank you for contacting the Western University Office of Research Ethics (ORE) regarding an upcoming research study being undertaken between the University of Florida (UF) and yourself. As per our discussion, and further communication with Liliana Alverez, PhD (Postdoctoral Fellow & Sessional Instructor), this project aims to use secondary de-identified data for analysis by your team. Details of this study are included below:

UF Study Title: Outcomes of Driver Evaluation and Fitness-to-Drive screening, comparing characteristics of older adult drivers with and without Parkinson's Disease

1. Completely unidentifiable data-set containing the ratings and scores of an extensive data base of participants in the following driving assessment, web-based screening tools, and assessment tools: Useful field of view, visual tests, visual attention tests, cognitive tests, motor tests, on-road assessment scoring, the fitness to drive screening measure
2. Secondary analyses to be performed
   a) address where the UFOV "stacks up" compared to the other clinical tests of driving outcomes in drivers with Parkinson's disease and healthy controls
   b) determine the cut-off points for visual, visual-attention, cognitive, and motor performance measures found to be predictive of failing the on-road assessment;
   c) evaluate the relationship between the type and number of driving errors found to be predictive of failing an on-road assessment and the visual, visual-attention, cognitive, and motor performance measures also found to be predictive of failing an on-road test;
   d) determine how well the FTDS measure, when completed by drivers, caregivers and evaluators, predicts on-road outcomes for drivers with PD and healthy controls, using the on-road assessment as the goal standard; and
   e) develop and evaluate a short form of the FTDS

As per Chapter 2, Article 2.4 of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS-2, 2010), the ORE does not require a formal application since this study falls outside of the scope of the ORE and "relies exclusively on secondary use of anonymous information" that will be provided by the UF and "the process of data linkage or recording or dissemination of results does not generate identifiable information".

Please retain this letter for your records and note that Western University's Federal Wide Assurance number: FWA00000121. Please do not hesitate to contact me if you have any further questions

Thank you,

[redacted]

Western University, Research, Support Services Bldg., Rm. 5150
London, ON, Canada N6G 1G9 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/ethics

Appendix D

Methods

The University of Florida's (UF) Institutional Review Board approved the primary study and release of the de-identified data to researchers (PI Classen) at the University of Western Ontario (UWO) on July 22nd 2016. The UWO Office of Human Research Ethics certified this secondary data analysis as a board review exemption, in accordance with Chapter 2, Article 2.4 of the Canadian Tri-Council Policy Statement: Ethical Conduct for Research Involving Human (TCPS-2) on March 27, 2015. In the initial study, participants provided consent before joining the study. Participants also received compensation upon completion of the study.

Design

The research student used measurement theory to construct and validate a short form FTDS. Specifically for this study, de-identified-data previously collected by Classen and colleagues (2015) for the Psychometrics of the Fitness-to-Drive Screening Measure© study (hereafter, referred to as the primary study) were used.

Participants

Primary study. The study used a convenience sample of community-dwelling older drivers (*n* = 200) and their caregivers (family members, friends or formal caregivers; *n* = 200). Participants were recruited in North-Central Florida, United States, and Thunder Bay, Ontario, Canada. Recruitment strategies included advertisements in newspapers, distributing flyers locally, and word of mouth. Older drivers, between the ages of 65 and 85 years, were included if they had a valid driver's license, drove at the time of recruitment, and were cognitively and physically able to complete the FTDS and on-road assessment. Conversely, drivers were excluded if they were medically advised not to drive, experienced seizures, and/or took medication that impaired their central nervous system. Also, caregivers, between the ages of 18 and 85 years, were included if they observed the older drivers' driving behaviors in the past three months. They were excluded if they displayed signs (through telephone interview and observation) of physical or mental conditions that may have impaired their ability to complete the FTDS, or make valid observations about the older drivers' driving behaviors.

Secondary analysis. Only the 200 caregivers' FTDS responses, from the primary study, were included in the development and validation of the short form FTDS.

Procedure

Primary study. Older drivers first rated themselves on the FTDS. Then they completed a battery of clinical tests, followed by an on-road assessment. Each driver's caregiver also completed the FTDS.

Secondary study. With the caregiver FTDS responses from the primary study, this secondary analysis followed a three-step process to construct and validate the short form FTDS. The research team carried out: (1) an exploratory factor analysis, (2) item analysis, and (3) correlational analysis. The steps used will be further discussed in the data analysis section.

Data Collection and Management

The de-identified data received from UF was stored on a password-protected server network at UWO. The de-identified data was received in a SPSS database (Version 20; IBM Corporation, Armonk, NY). The data were converted to an encrypted excel file only accessible to the research team. Quality control was performed to ensure data completeness because data with missing responses cannot be used to run an exploratory factor analysis. Therefore, items that displayed a pattern of missing responses across participants were removed. For example, item 53 “Drive on a snow covered road” was not answered by more than a quarter of the participants. Upon closer observation it was noted that this specific item was not missing responses at-random. A pattern had emerged where drivers (probably from Florida) were not completing this item because they may have never driven in snow. Identifying similar patterns among other items, a total of nine items were excluded from the analysis, resulting in 45 of the 54 items used. The removed items (described in abbreviated form) were: 20. Drive on graded (unpaved) road; 26. Parallel park; 35. Drive in an unfamiliar urban area; 36. Control his or her car when going down a steep hill; 37. Exit an expressway or inter-state highway from the left-handed lane; 38. Drive in a highly complex situation (such as a large city with high-speed traffic, multiple highway interchanges and several signs); 40. Drive a different vehicle (such as another person’s car or a rental car); 53. Drive on a snow covered road; 54. Drive on an icy road. Subsequently, the student researcher removed the caregiver data that were incomplete. After removal the data had 190 completed FTDS caregiver response sets. These data were then further analyzed.

Data Analysis

All the data was analyzed using R software version 3.1.2 (R Core Team, 2015).

Exploratory factor analysis. Through exploratory factor analysis, the research team examined the number of underlying factors in the FTDS that contributed to the fitness to drive construct. First, the student researcher determined the *factorability of the dataset*. The research team used two measures of sampling adequacy (Bartlett’s test of sphericity and Kaiser-Meyer-Olkin (KMO)) to determine the dataset's factorability. Factorability determines whether the potential to extract distinct and unambiguous factors, that may contribute to the construct of fitness to drive, are present (Gorsuch, 1973; Yong & Pearce, 2013). Specifically, the student researcher used Bartlett's test of sphericity ($p < .05$) to test for patterns of correlations among the items. The student researcher used the KMO to calculate the proportion of variance that may be common amongst items, indicating the dataset's suitability (> .50) for EFA (Yong & Pearce, 2013). Factorability was inferred based on meeting the criteria for both Bartlett's test of sphericity and KMO. Next, the student researcher selected the *potential* factor models. The student researcher used scree plots and parallel analyses to determine the number of potential factor models. A scree plot analysis informed the researcher on the number of factor models to extract, by plotting a graph of the number of factors, against the dataset's eigenvalues. An eigenvalue is a measure of the variance that is accounted for by the specific factor. To determine the final number of factor models to select, the research team visually assessed the plotted eigenvalues above the inflection point. This point is the bend in the plotted line, indicating distinct factor models that can be further considered for analysis (Thompson, 2004). Since the inflection point on a scree plot may not always be distinct, making scree plots somewhat subjective, the student researcher also ran a parallel analysis test. Parallel analysis is similar to a scree plot, in that it plots the number of factors against the dataset's eigenvalues. However, in parallel analysis, multiple sets of different eigenvalues for randomly generated data are created and plotted on the graph as well. To determine the number of factor models to select for further analysis, the researcher visually assessed and chose the factor models whose eigenvalues (from the caregiver dataset) were plotted above the eigenvalues for the generated data (Brown, 2006; Thompson, 2004).

Next, the student researcher extracted the *factor structure* for each potential factor model. That is the items of the FTDS that make up each factor's structure within the selected factor model. As part of the extraction process, an oblique rotation with a Promax rotation technique was applied to allow factors to be correlated with one another (Yong & Pearce, 2013). This rotation method represents a more realistic relationship between factors, that can still indicate if factors are not correlated (Brown, 2009; Thompson, 2004; Yong & Pearce, 2013). Through this method the student researcher extracted each factor with the largest possible variance, meaning that the groups of items within each factor accounted for maximum variability, yielding the simplest and most optimal factor structures.

In addition to producing the factor structure for each potential factor model, the extraction process also produced corresponding factor loadings for each item. The factor loading is the degree of variance in the item determined by the factor. Factor loadings that did not meet the arbitrary value of .40 were removed (Brown, 2009). Furthermore, items were removed from the factor model if their variance was dispersed to more than one factor (DeVellis, 2006).

Finally, the research team selected *a factor model*. The research team examined each of the potential factor model's structures, the number of items within each factor, factor loading, and the total variance of the model. For every potential model, the research team examined the grouping of items within each factor structure for interpretable patterns. While the number of items within a factor was considered, factors with too little items (i.e., three items) were not considered, because a factor with few items would not accurately represent the factor's construct. The range of the factor loadings was examined for the items. Higher factor loadings were considered better than factor loadings close to or below .40. The total variance of each factor within each model was also considered, with higher total variances considered better than lower total variances. The research team suggested a factor model that best fit the *person, vehicle, and environment* domains for the FTDS (Classen et al., 2010), and labeled the suggested model's factor structure with titles that best fit the domains. As such, the research student labelled Factor 1 *Complex Driving Tasks* because all of its items represented complex driving situations requiring strategic planning and controlling of the vehicle in complex or challenging environments. Factor 2 was labelled *Routine Driving Tasks* because the items within this factor represented tasks requiring basic driving abilities to manipulate a motor vehicle successfully. Lastly, factor 3 was labelled *Visual Scanning* because the items within this factor consisted of tasks related to the visual scanning abilities of the driver.

Item analysis. To determine the extent to which the items in the selected factor model were reliable, the research team first quantified its *scale reliability* using Cronbach's alpha (α) and alpha-if-deleted and then quantified its *item reliability* using item-total-correlations (Portney & Watkins, 2009). For each factor, Cronbach's α was calculated to determine whether the items within the factor measured the same construct. A Cronbach's α value of .90 is considered high and reliable. Alpha-if-deleted was then calculated to examine the impact of alpha (homogeneity) on each factor through removing each item within the factor structure. If removing an item resulted in a value exceeding Cronbach's α, that item was removed from the analysis because when included it decreased the homogeneity of the factor structure (Krishnan, 2013).

Lastly, the researchers quantified *item reliability* using item-total correlations to test how each item relates to and predicts the overall scale (DeVellis, 2006; Portney & Watkins, 2009). Items were removed from the factor structure if the item-total correlation did not fall between .70-.90, reflecting moderate correlations, which result from strong internal consistency (Portney & Watkins, 2009). In fact, high item-total correlations (> .90) may be indicative of a redundant item and low item-total correlations (< .70) may be indicative of a different trait (Portney & Watkins, 2009).

Correlational analysis. Once the item analysis was completed, the student researcher determined the strength of the linear relationship between the existing FTDS and newly constructed short form by conducting a correlation analysis (Pearson's *r* or Spearman's rho, depending on normality of the data as per Shapiro-Wilk test). For Pearson's *r* or Spearman's rho a positive correlation of greater than .75 was considered to represent a good to excellent relationship (Portney & Watkins, 2009).

References

Brown, J. D. (2009). Choosing the right type of rotation in pca and efa. *JALT Testing & Evaluation SIG Newsletter*, *13*(3), 20-25.

Brown, T. A. (2006). *Confirmatory Factor Analysis for applied research*. New York, NY: Guilford press.

DeVellis, R. F. (2006). Classical test theory. *Medical Care*, *44*(11), 50-59. Retrieved from http://www.jstor.org/stable/41219505

Gorsuch, R. L. (1973). Using Bartlett's significance test to determine the number of factors to extract. *Educational and Psychological measurement*, *33*, 361-364.

IBM Corp. 2011. *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Krishnan, V. (2013). The early child development instrument (EDI): An item analysis using classical test theory (CTT) on Alberta's data. *Early Child Development Mapping (ECMap) Project.*

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical practice: applications to practice*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall Health.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

Curriculum Vitae

Shabnam Medhizadah, BSc

Education

Oct 2013	**Honours Bachelor of Science, Psychology** *York University*, Toronto, ON

Scholarships, Honours, and Awards

2016	**Semi-finalist for the 2016 Geneva Challenge** *The Graduate Institute Geneva,* Geneva
2015	**Queen Elizabeth II Diamond Jubilee Scholar** *The University of Western Ontario*, London, ON
2015-2016 2014-2015	**Western University Graduate Research Scholarship** *The University of Western Ontario*, London, ON
2011-2013	**Member of Deans Honour Roll** *York University*, Toronto, ON
2008	**York University Entrance Scholarship** *York University*, Toronto, ON

Publications

1. Classen, S., & **Medhizadah, S**. (2016). Fitness-to-Drive Screening Measure: A valid and reliable tool for occupational therapy practice. *SIS Quarterly Practice Connections*. In press.

2. Knott, M., Alvarez, L., Krasniuk, S., & **Medhizadah**, **S**. (2016). Writing for The Reader's Brain: A Book review. *The Occupational Therapy Journal of Research: Occupation, Participation and Health, 36*(2), 99.

3. Classen, S., **Medhizadah, S.,** & Alvarez, L. (2016). The Fitness-to-Drive Screening Measure©: Patterns and trends for Canadian users. *The Open Journal of Occupational Therapy, 4(*4*), 1-15.*

4. Classen, S., **Medhizadah, S**., & Alvarez, L. (2015). The Fitness-to-Drive Screening Measure: Enabling caregivers to detect at-risk older drivers. *Occupational Therapy Now, 17*(5), 12.

Conferences

1. Alvarez, L., Classen, S., & **Medhizadah, S.** Efficacy of Driving App Research for Empowering (i-DARE) Teen Drivers. Poster to be presented at the Canadian Association of Occupational Therapy Conference 2016, Banff, Alberta, Canada.

2. Classen, S., **Medhizadah, S.,** & Alvarez, L. The Usability of the Fitness-to-Drive Screening measure within the Canadian Context. Poster presented at the Association for Driver Rehabilitation Specialists Conference 2015, Louisville, Kentucky, United States of America.

3. Classen, S., & **Medhizadah, S**. Fitness-to-Drive Screening Measure: Patterns and Trends for Canadian Users. Poster presented at the Faculty of Health Sciences Conference 2015, Western University, London, Ontario, Canada.

Employment and Internships

Jan 2016- Dec 2016 **Graduate Research Assistant for the i-DARE project**
i-Mobile Research Lab, The University of Western Ontario, London, ON

Sept 2015-Dec 2015 **Teaching Assistant**
The University of Western Ontario, London, ON

May 2015-Aug 2015 **Research Intern**
Catholic University of Health and Allied Sciences, Mwanza, Tanzania

Jan 2015-April 2015 **Teaching Assistant**
The University of Western Ontario, London, ON