Final Report
to the
Florida Department of Transportation

Project Number: **BDK77 977-17**

Project Title: **Validity and Usability of a Safe Driving Behaviors Measure for Older Adults**

for period 4/2011 to 4/2012

From: **Dr**. **Sherrilene Classen**, Department of Occupational Therapy, 101 South Newell Drive, Gainesville, Florida, phone (352) 273-6062, e-mail sclassen@phhp.ufl.edu

Date prepared 10/2012

DISCLAIMER PAGE

TECHNICAL REPORT DOCUMENTATION

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Validity and Usability of a Safe Driving Behavior Measure for Older Adults		5. Report Date October 15, 2012
		6. Performing Organization Code
7. Author(s) Classen, S., Velozo, C., Winter, S., Wang, Y. & Lanford, D.		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Florida - Department of Occupational Therapy PO Box 100164 Gainesville FL 32610		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. BDK77 977-17
12. Sponsoring Agency Name and Address Florida Department of Transportation 605 Suwannee Street, MS 30 Tallahassee, FL 32399		13. Type of Report and Period Covered Final report for period 04/11/2011 to 04/15/2012
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Provide a brief description of the problem, the objectives, the findings, the conclusions, and the benefits

With the aging of the Baby Boomers and ensuing Gray Tsunami in Florida leading the USA, older drivers who are unfit to drive must be identified. The gold standard on-road test is expensive, sophisticated, not available to many older drivers, and can only be validly conducted by a certified driver rehabilitation specialist of whom we have about 500 in the nation. To overcome these limitations, we have developed the Fitness-to-Drive Screening Measure (FTDS, previously known as the Safe Driving Behavior Measure or SDBM)**,** a free web-based tool ready for use by certified driver rehabilitation specialists (CDRS), occupational therapy practitioners, other health care professionals, and the caregivers/ family members of older drivers. This study used mixed methods, including item response theory, to refine self-report and proxy versions of a screening tool measuring driving behaviors of older adults, the SDBM. Understanding a driver's level of ability or difficulty with driving behaviors is a critical step providing a logical entry point for effective interventions, identifying optimal training parameters, and predicting future driving ability. We determined measurement properties for the SDBM, including validity (face, content, construct and criterion validity), factor structure, dimensionality, and item/person-level psychometrics. We determined the rater severity of the three rater groups (older driver, caregiver, and driving evaluator) and the criterion validity of the SDBM in relation to the reputed gold standard, on-road driving evaluation. Lastly, we developed the SDBM as a Web-based tool, including keyforms (rating profiles), ratings-based driver categories (i.e., basic, routine, and accomplished), each with targeted recommendations (i.e., next steps to assist in determining fitness to drive or to support driver health, knowledge, and skills). The findings suggest that this measure may be useful for: (1) family members/ caregivers to identify at-risk older drivers and to follow logical next steps based on keyform recommendations; (2) occupational therapy practitioners to identify an entry point for further interventions or referrals; and (3) CDRS to develop realistic and targeted intervention

goals to promote driving fitness. However, although not empirically tested, we also believe that this tool, available on-line: http://fitnesstodrive.phhp.ufl.edu, will have applicability to be used or disseminated among other health care practitioners, agents of the aging network, and stakeholders of the Safe Mobility for Life Coalition, in Florida.			
17. Key Word elder road users, aging drivers, screening, safe mobility	18. Distribution Statement		
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages 83	22. Price

EXECUTIVE SUMMARY

Background: Crash/injury/fatality rates of drivers over the age of 75 may increase with the future growth of this population. Accurate measurement of older-driver behaviors is imperative to curtail adverse effects of unsafe driving. The Comprehensive Driving Evaluation (CDE) consists of clinical tests related to driving (vision, cognition, motor, and sensory) in combination with an on-road evaluation of driving, usually conducted by an occupational therapist who is a certified driver rehabilitation specialist or CDRS. The CDE is a valid, reliable test, often referred to among driving behavior researchers as the gold standard, but it can be difficult to get older drivers to participate because of time needed, limited access, and threat of license loss. Alternatively, self-report is a means to identify older adults' safe driving behaviors, increase safety awareness/ knowledge, and promote behavior change and safer driving outcomes. We found existing measures fell short of providing meaningful descriptions of driving ability and did not contribute to risk reduction or increasing driving safety strategies. Understanding a driver's level of driving behaviors is a critical step towards providing an entry point for logical and effective interventions, identifying optimal training parameters, and predicting future driving ability. Towards this aim, we employed item response theory (IRT) methods to improve measurement precision of the Safe Driving Behavior Measure (SDBM).

Objectives: Our objectives were (1) to establish the psychometrics of the SDBM, including (a) factor structure, (b) item and person level properties and (c) rater severity; (2) to validate the SDBM to the reputed gold standard, on-road driving evaluation; (3) to develop the instructional clinical outputs, or ―keyforms (results summary of SDBM) and subsequently determine if driving evaluators, occupational therapists, older drivers and family members/caregivers can understand and interpret the results of the SDBM; and (4) to develop a data collection system capturing information from the end-users (older drivers, family members/caregivers) for further refinement of the SDBM, provide users with access to retrieve keyforms with recommendations, and track SDBM access and use.

Results: Psychometrics were established as follows. Through focus groups, we have established *face validity.* We established *content validity,* achieving 84% on a content validity index completed by expert reviewers. We determined the *construct validity* via Rasch analysis, identifying the person-and-item fit hierarchy of the items, structure of the rating scale, and homogeneity of the construct (fitness to drive). We determined the *construct validity* of the SDBM via Rasch analysis and identified the person fit and item fit. We have determined *unidimensionality* with factor analysis as well as *rater reliability* among three rater groups (older drivers, family members/caregivers, and the driving evaluator), and *rater effects* (level of leniency or severity) among these groups. We have showed through *concurrent criterion validity* that the older drivers showed statistically significant, yet poor, concurrent criterion validity compared to the family members/caregivers, who showed good concurrent criterion validity to the on-road driving test. We tested the usability, appearance, and acceptance of the Web-based SDBM through focus groups with occupational therapists, certified driver rehabilitation specialists, and family members/ caregivers. Lastly, we developed a keyform, or visual output summary of the caregiver ratings. Based on their ratings, this output summary (i) classifies a driver in one of four main groups; (ii) provides personalized examples of real world driving challenges; (iii) recommends logical next steps for the caregiver; and (iv) suggests general health and fit-to-drive strategies.

Conclusion: The findings suggest that the SDBM may be useful for: (1) family members/ caregivers to identify at-risk older drivers and to follow logical next steps based on keyform recommendations; (2) occupational therapy practitioners to identify an entry point for further interventions or referrals; and (3) certified driver rehabilitation specialists to develop realistic and targeted intervention goals to promote driving fitness. Although not empirically tested, we also believe that this tool, available on-line: http://fitnesstodrive.phhp.ufl.edu, will have applicability to be used or disseminated among other health care practitioners, agents of the aging network, and stakeholders of the Safe Mobility for Life Coalition, in Florida.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AOTA - American Occupational Therapy Association
AUC - area under the curve
CDE - comprehensive driving evaluation
CDRS - Certified Driver Rehabilitation Specialist
CFA - confirmatory factor analysis
CFI - comparative fit indices
CI - confidence interval
OT- occupational therapist
D - driver
E - evaluator
F/C - family member / caregiver
FTDS - Fitness-to-Drive Screening Measure
ICC - intra-class correlation
IRT - Item Response Theory
M - mean
MnSq - mean squared
NPV - negative predictive value
PPV - positive predictive value
RMSEA - root mean square error of approximations
ROC - receiver operating characteristic
SD - standard deviation
SDBM - Safe Driving Behavior Measure
TLI - Tucker-Lewis indices
VAS - visual analogue scale

CHAPTER 1 - INTRODUCTION

With the aging of the Baby Boomers and ensuing Gray Tsunami, and in Florida leading the nation in older driver demographics, older drivers who are unfit to drive must be identified. Crash/injury/fatality rates of drivers over the age of 75 will increase with the future growth of this population. Accurate measurement of older driver behaviors is imperative to curtail adverse effects of unsafe driving. The Comprehensive Driving Evaluation (reputed gold standard test) is highly valid and reliable, but limitations include being time-consuming, providing limited access, and holding an element of threat (mandatory/ ethical reporting upon failing). Self-report is a means to identify older adults' safe driving behaviors, increase driving safety awareness/ knowledge, and promote behavior change and safer driving outcomes. Existing measures are limited in accurately assessing older driver behaviors due to length, respondent burden, and inadequately representing driving constructs (i.e., person, vehicle, and environment). Current self-report measures fall short of providing meaningful descriptions of driving ability level, and do not contribute to targeting risk reduction or increasing driving safety strategies. In contrast, item response theory (IRT) or Rasch analysis methods are particularly useful to measure driving behaviors. IRT addressed both the difficulty of a behavior (or item) and person ability to understanding a driver's "level" of driving behaviors. Such knowledge is a critical step towards providing an entry point for logical and effective interventions, identifying optimal training parameters, and predicting future driving ability. To that end we proposed further development of the SDBM.

Research Tasks

Objective 1: Establish the psychometrics of the SDBM
Task 1(a): Confirm the factor structure of the SDBM with a sample of 200 dyads (drivers and family members/caregivers [F/C]).
Task 1(a) Hypothesis: The SDBM has a two-factor structure: pre-driving skills and driving skills.
Task 1(b): Describe the item and person level properties of the SDBM based on findings from a sample of 200 dyads (drivers and F/C).
Task 1(b) Hypothesis: The SDBM will fit the Rasch measurement model.
Task 1(c): Determine the rater severity of the three rater groups (older driver, F/C, driving evaluator) based on findings from a sample of 200 dyads (drivers and F/C).
Task 1(c) Hypothesis: The evaluator will be the most severe rater, followed by the F/C, and then the older driver.

Objective 2: Validate the SDBM to results from on-road driving evaluation.
Task 2: Conduct a concurrent criterion validity study using the SDBM and on-road test results from a sample of 200 dyads (drivers and F/C).
Task 2 Hypothesis: The SDBM will be a valid instrument of predicting passing/failing an on road driving test.

Objective 3: Develop the instructional clinical outputs — or keyforms — to determine if driving evaluators and occupational therapists (OTs) as well as older drivers and F/C understand the results of the SDBM in an interpretable way.

Task 3(a): Using focus group methodology, determine if driving evaluators and occupational therapists (OTs) understand the results of the SDBM in an interpretable way.

Task 3(a) Hypothesis: The SDBM will be used by clinicians as a clinically useful measure to discern the level of safe driving ability of the older driver and to identify the next logical steps for intervention to improve safe driving behaviors.

Task 3(b): Using focus group methodology, determine if older drivers and F/C understand the results of the SDBM in an interpretable way.

Task 3(b) Hypothesis: The SDBM will be used by older drivers and F/C as a useful measure to discern the level of safe driving ability of the older driver and to identify the next logical steps for intervention to improve safe driving behaviors.

Objective 4: Develop a data collection system that will capture the SDBM information from the end-users (older drivers and F/C) for further refinement of the SDBM, provide users with access to retrieve keyforms with recommendations, and track the access and use of the SDBM via an electronic tracking system.

Task 4 Hypotheses: None – this Objective is an application of Objectives 1-3.

Task 4(a) Capture the SDBM information from the end-users (older drivers and F/C) for further refinement of the SDBM.

Task 4(b) Provide older drivers and F/C with access to retrieve keyforms from the identified websites such as I-MAP; SafeandMobileSeniors.org, and the American Occupational Therapy Association (AOTA) with recommendations regarding continued safe driving, driving with restrictions, referrals, or driving cessation.

Task 4(c) Track the access and use of the SDBM via an electronic tracking system.

CHAPTER 2 - BACKGROUND

This chapter provides the theoretical background, rationale, and review of relevant literature for objectives 1 through 4.

Objective 1: Establish the psychometrics of the SDBM.

Task 1(a): Confirm the factor structure of the SDBM with a sample of 200 dyads (drivers and F/C).
An underlying assumption of Rasch analysis or Item Response Theory in development of a measure, is that one is measuring a unidimensional construct. That means all the items are working together to measure a single characteristic. From our preliminary work, we anticipated that the 68 items of the SDBM represented two dimensions, pre-driving skills and driving skills. For this reason, a confirmatory factor analysis was planned to inform us about the dimensionality of the items. This knowledge was used to design the items of the measure accordingly, for example by separating items into sections based on the dimension represented.

Task 1(b): Describe the item and person level properties of the SDBM.
Focusing on the strengths of self and proxy report measures, we developed the Safe Driving Behavior Measure (SDBM), and research findings reflected the face and content validity, rater reliability, and rater effects (Classen et al., 2010; Winter et al., 2011; Classen et al., 2012a; Classen et al., 2012b). The SDBM consists of three sections: A = Demographic information; B = Driving habits; C = Driving behavior questionnaire with 68 items (sample items in Appendix A) and a proposed hierarchy of driving tasks which increase in complexity. For example, the instrument indicates that item 1 "open car door" is potentially the easiest item and that item 68 "drive on an icy road" is potentially the most difficult item. Based on this principle, one may assume that if a person can drive in "an unfamiliar urban area" without difficulty (item 49), then he/she may also have a high probability to be successful with the preceding items. Understanding the "level" of safe driving behavior of a participant is a critical step towards providing an entry point for occupational therapists to plan logical and effective interventions, identify optimal training parameters, and to predict future safe driving ability.

The objective of this project was to investigate the item/person-level psychometrics and item hierarchy of three groups: older drivers, F/C, and driving evaluators who had completed the 68-item SDBM. If the SDBM shows reasonable psychometric properties, it will assist occupational therapy generalists to identify unsafe driving behaviors and provide them with an entry point for delivering preventative services. Psychometrics of the SDBM for this study included i**tem statistics** (i.e., *item difficulty, item fit, item reliability,* and *item separation*), **person statistics** (i.e., *person's ability, person fit, ceiling/floor effects, person reliability,* and *person separation*), and **item hierarchy**. *Item difficulty* is an estimate of an item's underlying difficulty calibrated from the total numbers of drivers who succeed on the item. *Item fit* was determined by the fit statistics of each item provided by the Winsteps program. The Winsteps program provides two types of fit statistics: information-weighted mean square (infit MnSq) and outlier-sensitive mean square (outfit MnSq). The ratings of a driver that a rater assigned in the highest and lowest categories of the scale are weighted less heavily on the Infit MnSq. The infit MnSq has an expected value of 1. Values > 1 signal more variation (i.e., unexplained, unmodeled variation) in a driver's ratings on the items than expected; values < 1 signal less variation in a driver's ratings on the items than expected. Generally, infit > 1 is more of a problem than infit < 1, since highly surprising or unexpected ratings that do not "fit" with the other ratings tend to be more difficult to explain and defend than overly predictable ratings. By contrast, the outfit MnSq is more sensitive than the infit MnSq statistic to the occasional highly unexpected and surprising ratings that may occur; therefore we used infit statistics. The criteria of the Infit MnSqs were set from criterion (A) 0.6 – 1.4, and criterion (B) 0.5 - 1.7 and the standardized fit statistics were set from -2 to 2, (Type 1 error rate = 0.05) (Wang & Chen, 2005; Wright & Linacre, 1994). *Item reliability* represents how well the estimates of the item measures can be replicated, when another sample with comparable ability are rated, using the same set of items. *Item separation* estimates how well the items are separated by the measured variable.

Person's ability is an estimate of the driver's underlying ability based on the driver's performance on a set of items and person ability is calibrated from the total number of items to which the driver responded successfully. Similar to item fit, *Person fit* is determined by the fit statistics of persons, with person misfit indicating that one or more of the ratings that the rater assigned to the older driver were surprising or unexpected. *Ceiling effect* is defined as more than 5% of participants rated at the maximal score, while *floor effect* is defined as more than 5% of participants rated at the minimal score. *Person reliability* represents how well the estimate of the driver's ability can be replicated when other sets of items, measuring the same construct, are used to rate the same sample of drivers, and is analogous to Cronbach's alpha with values between 0 and 1. *Person separation* index, measured in standard error units, indicates how well the instrument separates drivers of different levels of safe driving ability. The statistically distinct strata of safe driving ability within the sample of older drivers can be obtained by applying the formula $(4G_p+1)/3$, where G_p represents the person separation index (Wright & Masters, 1982). An assessment needs at least two strata to reliably distinguish between safe and unsafe older drivers.

Item hierarchy was evaluated based on the item map provided by the Winsteps program. One of the strengths of the Rasch model is that it can readily handle missing data or "Not applicable" answers. That is, the Rasch model does not require a fully crossed rating design (i.e., a design which requires all raters to rate all items); it can easily accommodate partially crossed rating designs that provide sufficient linkage of raters and drivers.

Task 1(c): Determine the rater severity of the three rater groups (older driver, F/C, driving evaluator). In this study we address inter-rater reliability among three groups of raters (older driver, F/C and driving evaluators), and investigated the rater effects among two of the groups (F/C and driving evaluators) on the SDBM. We expect that the findings of this study will provide the evidence to use the self/ proxy-report SDBM as a reliable measure of safe driving among older adults, their F/C, and occupational therapists who are not trained Certified Driver Rehabilitation Specialists (CDRSs).

In previous work we have developed a self-report Safe Driving Behavior Measure (SDBM) (Classen et al., 2010; Winter et al., 2011). In this project, as tested among 200 older drivers, 200 F/C and two driving evaluators we examine the SDBM as a precise and accurate measure for detecting safe driving behaviors among older adults.

Inter- rater reliability

Inter-rater reliability is defined as the extent to which different raters agree on the same persons or characteristics. The terms, inter-rater reliability, rater agreement, and rater correlation are often used interchangeably. Two studies investigated the relationship of driving performance rated by evaluators and older drivers (Marottoli & Richardson, 1998; Wild & Cotrell, 2003). Marottoli and Richardson investigated the relationship between on-road driving performance rated by evaluators and self-reported driving ability rated by older drivers using the Pearson correlation. They did not find a significant association between the ratings of these two groups. Wild and Cotrell investigated the differences between evaluators' ratings and older drivers' ratings on the Driving Safety Evaluation scale using t- test statistic. They found that only 2 of 10 items showed significant differences between the evaluators' ratings and the older drivers' ratings.

Neither the Pearson correlation coefficient nor t-test statistic, can accurately determine the potential rater effects. Rater effects are influences on the ratings attributed to characteristics of the raters. Rater effects of particular interest are accuracy vs. inaccuracy, severity vs. leniency, and centrality vs. extremism of scores (Myford & Wolfe, 2003). While the Pearson correlation can provide the strength of the relationship between two sets of data (the concordance of the data), it cannot detect if the value of one set of data is consistently greater than the other one, which might indicate that one rater is more severe or lenient than the other. The t-test statistic detects the significant difference of the means of two sets of data; however, using the "mean" may partial out the individual differences that exist within the rater group. Further, the Pearson correlation and t-test statistic cannot provide information regarding the response pattern; that is, whether someone rates inconsistently. Thus, although the Pearson correlation and t-test statistics provide necessary methods for assessing rater agreement, it is not sufficient to make an accurate determination of rater effects. It is critical to examine rater effects, especially when individuals will be reporting on safety aspects of driving.

Rater Effects

Rater effects is a function of *severity or leniency* defined as the tendency that a rater assign ratings consistently higher or lower than other raters (Myford & Wolfe, 2004). In addition to having tendency to assign higher or lower ratings, raters may also assign ratings erratically (erratic response pattern); that is, the raters assign, inappropriately, low scores (cannot do) to drivers with a higher ability level, or high scores (no difficulty) to drivers of a lower ability level. The Many Facet Rasch Model (MFRM), an extension of the Rasch model, is useful to investigate the rater severity and response patterns (Linacre, 2004). The Rasch model, a one-parameter model IRT converts ordinal scales into interval measures (using logit as its unit) and provides a useful, efficient and objective framework for developing, evaluating, and revising measures. Recently, five published driving studies applied Rasch analysis to develop or evaluate driving scales (Kay, Bundy, Clemson, & Jolly, 2008; Kay et al., 2009; Myers et al., 2008; Patomella et al., 2008; Patomella et al., 2006). Patomella and colleagues (2006) first applied Rasch analysis to examine the Performance Analysis of Driving Ability (P-Drive), in a driving simulator, with 31 persons with brain injury; and later (2008) they used Rasch analysis to evaluate the P-Drive with 101 individuals with stroke. Another study (Kay, Bundy, Clemson, & Jolly, 2008) applied Rasch analysis on a standard on-road test to transform the on-road test into a linear interval measure with hierarchical ordered tasks. Myers and colleagues (2008) examined the structure of a scale assessing driving confidence using Rasch model. Most recently, Kay and colleagues applied Rasch analysis on a simulated test rated by trained professionals, and an awareness test, to investigate the construct validity and internal reliability of the two scales (Kay, Bundy, Clemson & Jolly, 2008). While we are seeing an increased application of Rasch analysis in developing and evaluating assessments, no driving-related published study has yet applied the Rasch model to assess rater effects.

Beyond estimating item difficulties and person abilities, the Many Facet Rasch Model (MFRM) includes an additional parameter, the rater to detect whether the response differences are caused by systematic rater severity/ leniency. Moreover, by fitting data to the Rasch model, the MFRM can detect the erratic raters. Rater effects are important to compare driver vs. proxy and evaluator ratings. When comparing older driver self-reports with F/C reports or driver evaluator reports, we anticipate a discrepancy. That is, we expect that older drivers may be the least severe in their self-ratings (e.g., not wanting to lose their license) and the evaluators may be the most severe in their ratings (e.g., their training is to focus on deficits); the F/C may be somewhat in the middle with their ratings as some may be overly severe (i.e., really want the driver to stop driving) and some less severe (i.e., not wanting to lose their means of transportation) with the driving safety of their loved one.

Objective 2. We validated the SDBM to results from on-road driving evaluation.
The SDBM has promise to be used as a screening tool for F/C and potentially the older drivers, but the criterion validity had to be established against the on-road test conducted by trained driving evaluators.

Existing valid self or proxy-reports tested against the reputed on-road gold standard measure are limited in the driving literature. Caregiver opinions have been sought in several driving studies. For example, Wild and Cotrell (2003) found that caregivers had insight into the driving errors (e.g., managing intersections, managing lane changes) of care recipients with Alzheimer's disease who still drove. However, they underreported some driving errors of the care recipients when compared to a standardized road test. Croston et al., (2009) reported that family members could provide adequate information on some driving behaviors (e.g., monitoring traffic, maintaining speed) of drivers with dementia (Alzheimer's type). In our previous work, we found that F/C were more reliable than healthy community-dwelling licensed drivers to report on driving behaviors but they were not as accurate as the driving evaluator reports which were based on standardized on-road tests (Classen et al., 2012b, Classen et al., in press).

Recognizing that caregivers make an important contribution in identifying driving errors or driving behaviors, we have used their input in determining the psychometrics of the SDBM. As such, F/C were involved in establishing face and content validity (Classen et al., 2010; Winter et al., 2011); and their ratings were used to determine construct validity (Classen et al., 2012a), rater reliability, and rater effects (leniency vs. severity) among three rater groups (older drivers, F/C, driving evaluators) (Classen et al., 2012b). Our preliminary findings (from the cited studies above) pointed to the usefulness of the SDBM as a screening measure used by F/C to rate the driving behaviors of older drivers, but needed testing among a larger more diverse group to test concurrent criterion validity.

Measure of validity testing: Receiver Operating Characteristic (ROC) Curves

Receiver operating characteristic (ROC) curves is a methodology to determine the criterion validity of a screening tool as measured against a gold-standard outcome. Essentially, the ROC curve is a plot of the rate of true positives (true hits or sensitivity) against the rate of false positives (true misses or 1-specificity) resulting from application of many arbitrarily chosen cutoff points of the predictor test (Portney & Watkins, 2000). The ROC curve demonstrates the effectiveness of using different cutoffvalues and reveals the optimal cutoff value for the predictor test. If the *area under the curve*, an index of discriminability, is statistically significant and at least 0.70 in magnitude, then further attention must be paid to the other ROC attributes, such as *sensitivity, specificity, positive predictive value* and *negative predictive value* (Portney & Watkins, 2000).

Sensitivity is the predictor test's ability to obtain a positive test when the condition really exists (a true positive); here it means that the predictor test would suggest the participant will fail the on-road test, and the participant actually fails it. *Specificity* is the predictor test's ability to obtain a negative result when the condition is really absent (a true negative), here the predictor test would suggest the participant will pass an on-road test, and the participant actually passes it (Portney & Watkins, 2000). *Positive predictive value (PPV)* is the probability that the participant will, given a certain cut-point on the predictor test suggesting a fail the on-road test, actually fail the on-road test. *Negative predictive value (NPV)* is the probability that the participant will, given a cut-point on the predictor test suggesting a pass for the on-road test, actually pass the on-road test. It is important to note that the number of *false positives* (those who receive a failing score but pass the road test), and *false negatives* (those who receive a passing score but fail the road test) and thus the sensitivity and specificity values, change with the cutoffvalue. Ultimately one wants the false positives and false negatives to be as close to 0 as possible. For an example of ROC curves using error scores to determine passing/ failing an on-road test see (Shechtman et al., 2009), and for using ROC to determine the sensitivity of predictor tests of on-road outcomes see (Classen et al., 2009).

Objective 3: Develop the instructional clinical outputs (result summaries), or "keyforms" to determine if driving evaluators, OTs and F/Cs understand the results of the SDBM in an interpretable way.

We used contributions of three stakeholder groups: OTs, expert CDRSs, and F/C. Specifically, we conducted three focus groups, one with each stakeholder group, to learn their needs, perspectives, and suggestions for refining the Web-based SDBM and keyform. Our research question was: *What is the input (needs, perspectives, and suggestions) of stakeholders (occupational therapy practitioners, CDRSs, and family members/caregivers) in the process of developing a "keyform" for the SDBM?*

Stakeholders, i.e., occupational therapy practitioners, expert raters, or F/C may contribute to developing measures for clients. For example, the OT may acquire knowledge on habits of a driver during the interview (American Occupational Therapy Association, 2010); an expert rater, such as a CDRS may interpret the drivers' on-road performance to reveal errors or violations (Classen et al., 2010); or a family member riding with the driver may observe lapses or near misses (Wild & Cotrell, 2003). In fact, Thurstone (1925) suggested that the development of a measure starts by understanding the qualitative experiences shared by the stakeholders. Such qualitative experiences are best captured by soliciting the perceptions of stakeholders who have real life experiences with the client (Magasi & Heinemann, 2009). The qualitative features of the stakeholder-client interaction therefore constitutes a real-time interpretation of the driver's abilities, and subsequent formulation of the needs, perceptions or suggestions pertaining to the driver in his or her context. This interaction, between stakeholder and client, is essential to capture in the development of outcome measures.

Driving is a key area needing to be addressed by all OTs (American Occupational Therapy Association, 2010). Both generalists and those who specialize in the assessment and rehabilitation of driving may contribute to measurement development. Failing to include these practitioners, with their understanding of the background, clinical utility and application of the measurement tool, may lead to measures which lack essential information critical to clinical decision making. Thus, without practitioner input, a gap may continue to exist between measurement tool development and the *translation of the measurement tool* to clinical practice.

Experts are recognized persons whose skill, knowledge, or judgment in a specific well-distinguished domain is widely recognized by their peers and/or the public. In the practice area of driving, persons designated as either Certified Driver Rehabilitation Specialist (CDRS) or having the Specialty Certification in Driving and Community Mobility (SCDCM) are considered experts (American Occupational Therapy Association, 2010). Given their extended involvement in assessing driving and remediating driving performance issues (and community mobility), CDRSs and SCDCMs can make important contributions to measure development and refinement in at least three ways. First, by nature of their prolonged or intense experience through practice and education they bring depth to the meaning of items measuring driving performance. Next, they may identify gaps or particular strengths in the measurement instrument. Finally, using their clinical reasoning and critical thinking skills, they can help with the interpretation of data and make useful recommendations for intervention. Although experts in occupational therapy are used widely to provide "expert witness testimonies", they are not regularly included in the development of measurement tools. Such inclusion can bring domain and content specific knowledge not otherwise obtained through contributions of generalists.

Caregivers have been used as a group to assess driving errors or driving behaviors of their loved ones or care recipients (Wild & Cotrell, 2003). They also have been included in the development of driving measures. For example, in our prior work with the Safe Driving Behavior Measure (SDBM), we have involved F/C in: establishing face and content validity (Classen et al., 2010; Winter et al., 2011); determining construct validity (Classen et al., 2012a); determining rater reliability and rater effects (leniency vs. severity) among three rater groups (older drivers, F/C, driving evaluators) (Classen et al., 2012b); and determining criterion validity of the SDBM to on-road outcomes (Classen et al., in press). In fact, we found that F/C were more reliable than healthy licensed older drivers to report on driving behaviors, yet not as accurate as the CDRS reports which were based on standardized on-road tests (Classen et al., 2012b). The F/C also demonstrated acceptable accuracy in their SDBM ratings to predict on-road outcomes among older drivers (Classen et al., in press).

A keyform is a results summary of the clinical assessment that illustrates the relationship of client performance on the items of an instrument. This form is generated from the “General Keyforms” output table produced from Rasch analysis using the Winsteps software program (Linacre, 2012a; Linacre, 2012b). A core feature is that the keyform provides immediate and useful information to the stakeholder. For example, with a glance, the OT may observe the client’s profile, including tasks (expressed as items) that are “easy” to perform and tasks that are “hard” to perform. A major benefit of the keyform is that it provides an entry point for occupational therapy interventions (Kielhofner et al., 2005). However, by incorporating the perspectives of F/C in keyform development, it may provide an entry point for further family conversations, or decision making relevant to the independence or safety of their loved one.

Objective 4: Develop a data collection system that will capture the SDBM information from the end-users (older drivers and F/C) for further refinement of the SDBM, provide users with access to retrieve keyforms with recommendations, and track the access and use of the SDBM via an electronic tracking system.

As part of refining the SDBM, we developed a Web-based measure. Results for the Web-based measure were captured in a database maintained at the University of Florida within the College of Public Health and Health Professions. SDBM ratings provided via the Web-based format are scored electronically, providing a driver profile with a clinically relevant output indicating the type and severity of driving difficulty as well as safety recommendations. These outputs can be used by older adults, their F/Cs and OTs for decision-making. Suitable recommendations are provided after a screening has been completed via the Web-based method.

CHAPTER 3 - METHODS

Objective 1: Establish the psychometrics of the SDBM.

Task 1(a): Confirm the factor structure of the SDBM.

Design

This prospective quasi-experimental study used a convenience sample of 200 older drivers and their 200 F/C from two sites to examine the concurrent criterion validity of the SDBM against the outcome (pass/fail) of standardized on-road tests. This study received Institutional Review Board approval from the University of Florida (primary site) and a secondary site at Lakehead University, Thunder Bay, Ontario, Canada. All subjects who participated provided informed consent. We recruited older drivers and F/C from North Central Florida and Thunder Bay, Ontario, Canada, by flyer distribution in the local community facilities, local newspaper advertisements, and word-of-mouth referrals. Older drivers were included if they were 65-85 years of age, had a valid driver's license, were driving three months prior to and at the time of recruitment, and had the cognitive and physical ability to complete the SDBM and participate in an on-road driving test. They were excluded if they had: medical advice not to drive, uncontrolled seizures in last year; or used medications that cause central nervous system impairments. Family members/Caregivers (F/C) were included if they were able to report on the older adult's driving behaviors and excluded if they showed presence of physical or mental conditions that impaired their ability to participate.

Participants

We recruited participants in North Florida and Ontario, Canada, by advertisements in newspapers, word-of-mouth referrals, and flyers distributed to local community facilities (e.g., retirement communities) and a community-based clinic serving a rural older adults. A convenience sample of 200 older community-dwelling licensed drivers were included if they were between 65-85 years of age, had a valid driver's license, were driving at the time of recruitment, had the cognitive ability to complete the SDBM, and had the cognitive and physical ability to participate in an on-road driving test. Participants were excluded if they had been medically advised not to drive, experienced uncontrolled seizures in the last year, or took medications that caused central nervous system impairment. Eighty F/C (18-85 years of age) were included if they were able to report (based on observation) on the older adult's driving behavior and excluded if they showed presence of physical or mental conditions that impaired the ability to make an active contribution. At the primary site, the certified driving evaluator, an occupational therapist with six years of clinical practice experience, conducted the driving evaluations. At the Canadian site, the driving evaluator was an accredited driving instructor (Province of Ontario) and evaluator with over 10 years of experience. Thus, the rater groups were older drivers, F/C, and driving evaluators.

Procedure

All older drivers and F/C gave written informed consent before completing their demographic profiles

and the SDBM. Older drivers completed the SDBM questionnaire and our validated clinical battery (see Stav et al., 2008). In accordance with the protocol, drivers were next evaluated by trained driving evaluators via a standardized on-road driving evaluation (Justiss et al., 2006). The on-road driving test occurred on the same, or close to the same, day as the SDBM and clinical test administration, except if rain or adverse weather events interfered with the on-road test; in this situation, the on-road driving test was rescheduled on a different day. All aspects of testing were performed by a certified driver rehabilitation specialist (CDRS) at the main site, and by a trained driving evaluator at the Canada site. The evaluators had a 100% inter-rater reliability (Classen et al., 2010).The two evaluators (one per site) who were blinded to the participants' SDBM self-ratings or proxy ratings, also completed a SDBM on each driver after the on-road test. Drivers and F/C received $50 to $100 for their study participation.
Family members/caregivers completed SDBM questionnaire section A (Demographics) to provide information on themselves and their relationship with the driver (e.g., how often did they ride with the driver). They also completed section C (68 items of driving behaviors), based on their observations over the last three months.

Measures and Study Variables

Self-report and Proxy report

The SDBM is a 68-item self-report or proxy measure to assess safe driving behaviors (Classen et al., 2010; Winter et al., 2011). The measure score (derived from Rasch analysis) represents the reported level of difficulty for the items given the participant's ability level. Difficulty with the driving task is rated via a 5-point adjectival scale (from 1= Cannot do to 5=Not difficult). A copy of the SDBM is included with this report as appendix A.

Clinical tests

The validated clinical test battery, with reported psychometrics, included tests of vision, vision-cognition, cognition and motor performance and is fully documented in previous studies. For the purposes of this study we are only including information on the abilities described below (Stav, et al., 2008).
Vision. Visual acuity and contrast sensitivity were tested using the Optec® 2500 visual analyzer (Stereo Optical Company Inc., 2007). We categorized the binocular (both eye open) visual acuity as "20/20 to 20/40" and "20/50 or poorer (e.g., $\geq$20/70). We dichotomized contrast sensitivity as *intact* (all 5 Optec ® 2500 contrast sensitivity slides are intact) and *impaired* (any of the five contrast sensitivity slides is impaired).
Visual-cognition. We reported the Useful Field of View (UFOV) risk index (1 = very low risk, 2 = low risk, 3 = low-moderate risk, 4 = moderate-high risk, and 5 = high risk) and three UFOV subsets (UFOV 1 = visual search and visual processing; UFOV 2 = divided attention; and UFOV 3 = selective attention) (Ball & Owsley, 1993; Edwards, et al., 2006). The cut-point for each one of the sub-tests is 500 milliseconds, meaning that if a person exceeds this score per subtest, he/she will not be able to continue the proceeding sections and may have impaired visual processing speed.

Cognition. We used Mini Mental State Examination (MMSE, maximum score = 30) as an indicator of baseline cognitive functioning (Folstein et al., 1975).
Motor performance. We used the Rapid Pace Walk (RPW) (in seconds) to test the motor performance (gait, postural control, balance and speed of walking) of older drivers. The RPW when executed for longer than 7 seconds is predictive of adverse driving events (accidents, violations, being stopped by the police, violation, or traffic accident) (Marottoli et al., 1994); and this test is statistically significantly correlated to on-road driving performance (Stav et al., 2008).

On-road test
The Florida on-road tests consisted of driving a standardized road course with demonstrated reliability (ICC= 0.94, $p<0.05$) and validity (driving performance score was correlated to the Global rating score, $r=0.84$, $p<0.001$) for older drivers (Justiss et al., 2006; Posse et al., 2006; Bédard et al., 2008). The Canadian site used a demerit point system consistent with the method used by their licensing authority. The outcome of the road-tests included a pass/fail measure of driving: 3 = Pass, 2 = Pass with restrictions or recommendations, 1 = Fail with remediation, 0 = Fail not remediable. Both UF and Lakehead used a dichotomized pass/fail outcome.

Data Collection and Management
All the data (SDBM, demographic information, scores of the clinical tests and the on-road tests) of the older drivers and F/C were collected and recorded by research assistants in a central secure and password protected data repository, which was located at the primary site, at the University of Florida. Data entry was monitored by the principal investigator and quality control spot checks and corrections were made, intermittently, to ensure data completion and accuracy.

Data Analysis
First, we examined the frequency of each rating scale category to see whether the rating categories were adequately used. For example, if items are too easy, most of raters will use a 5 (Not difficult). Because of the lack of variance, these items cannot distinguish between a driver's ability levels. Therefore, we excluded items when ≥ 95% of the ratings on the items represented a "5" (Not difficult). However, we also weighed "losing" items that may contribute to the concept of safe driving, even if those items did not meet the 95% standard mentioned above. Additionally, we found that rating category 1 (Cannot do) and 2 (Very Difficult) were underutilized by all three rater groups; i.e., per rater group, raters utilized 1 and 2 only 2% among all the category ratings. Based on this underutilization pattern we collapsed these categories under the rating of "1" (Very difficult).

Task 1(b): Describe the item and person level properties of the SDBM

Details of design, participants, procedure, measurement and data collections are listed above under Objective 1(see pages 10-12).

Data Analysis

We managed the participant demographic data using SPSS (Version 20) (SPSS Inc., 2012), and we used the rating scale model implemented through the Winsteps computer program (Version 3.74.0) (Linacre, 2012a) to conduct Rasch analyses of the rating data. In using the rating scale model, we assumed that the rating scale structure was similar across the 54 items on our instrument. That is, we assumed that the raters used each of the categories of the rating scale in a similar fashion when rating each item (i.e., a "1" on item 1 was equivalent to a "1" on each of the other items; a "2" on item 1 was equivalent to a "2" on each of the other items, etc.).

We reported on the older drivers and F/C demographic information, and psychometric properties of the SDBM across the three rater groups. Rasch analysis is a one-parameter logistic model (1-PL) which assumes all items have a constant item discrimination parameter. Because of its simplicity, the Rasch model, unlike 2-PL or 3-PL models, does not require large sample sizes to obtain stable estimates and is preferred in the rehabilitation field (Jette & Haley, 2005). For polytomous scales (such as the SDBM), the rating scale model of Rasch analysis that calibrates the rating scale across all items using the same rating scale structure, is a preferable model for small samples; hence adequate to perform data analyses on our sample (N=200 for each rater group) (Linacre, 2000). The measurement model we employed is presented below:

$$\text{Log}\,[P_{nik} / P_{ni(k-1)}] = B_n - D_i - F_k$$

Where P_{nik} = probability of driver n receiving a rating of k on item i,

$P_{ni(k-1)}$ = the probability of driver n receiving a rating of k-1 on item i

D_i = the difficulty of item i, and

F_k = the difficulty of receiving a rating of k, relative to receiving a rating of k-1.

The rating scale structure was investigated according to three essential criteria: 1) at least 10 observations per rating category; 2): the average measures (mean of each category) should advance; that is if "cannot do" is -2 logits, the average measure of "a lot of difficulty" should be larger than -2 logits; and 3) outlier-sensitive mean square fit statistic for each rating scale category should be bigger than 2.0 (Linacre, 2002a). We collapsed rating categories when they were underutilized.

Task 1(c): Determine the rater severity of the three rater groups

Details of design, participants, procedure, measurement and data collections are listed above under Objective 1(see pages 10-12).

Data Analysis

Item inclusion/exclusion: We excluded 35 items from the analysis. These items included 21 items that were not observable by the driving evaluator at the time of testing (e.g., driving in snow), and 14 items that added little or no variance to the responses, or represented other construct. For example, > 95% of respondents used the same rating category, i.e., "not difficult" for 5 items. We used the collapsed 4 rating categories for the analysis.

Inter-rater reliability: We conducted an intra-class correlation (ICC) to examine the rater reliability on the 33 remaining items. We used SPSS version 20.0 (SPSS Inc., 2012) for the analyses and a p-value less or equal to 0.05 was considered significant for the correlations.

Rater effects: We conducted the Many-Facet Rasch Model (MFRM) to analyze rater effects using the Facets software version 3.70. (Linacre, 2012). The MFRM extends the rating scale Rasch model by adding one component/facet (Cj) to calibrate rater severity:

Log[Pnik /(Pni(k-1))]=Bn - Dgi - Fgk - Cj

Where Pnik =probability of observing category k for person n who answers item i;

Pni(k-1) =probability of observing category k-1;

Bn =person ability;

Dgi =item difficulty for item i in group g;

Fgk =the difficulty of being observed in category k relative to category k-1 for an item in group g; and

Cj= severity of judge j, who gives rating k to person n on item i

Facet ruler, fit statistics, fixed chi-square, and paired comparisons were used to investigate the rater effects. *Facet ruler*, displaying three Facets (rater, item difficulty, person ability) in the same linear continuum, provides a visual map to compare the relative hierarchy within and between Facets. To illustrate the relative distribution of the drivers' abilities and item difficulties simultaneously, we anchored the mean of the rater severity to 0. *Fit statistics* (Infit MnSqs and Outfit MnSqs) were used to detect erratic raters; that is, raters who assign high scores to drivers from a low ability level, and low scores to those drivers with a high ability level. Infit statistics is more responsive to the variance of those well-targeted observations, while outfit statistics is sensitive to the variance of outliers or extreme observations. Ideal fit is when the observed response patterns exactly match the predicted pattern (MnSq=1) of the model. Infit MnSq and Outfit MnSq ranging from 0.6 to 1.4 were considered adequate fit for survey data (Bond & Fox, 2001). The measure represents the average ratings of the rater in logits with higher scores indicative of greater severity in rating. *The fixed chi-square* was used to examine whether at least one rater group, on the overall scale level, consistently used the ratings differently from other rater groups. Should the fixed chi-square test be significant, then *paired comparisons* are performed to identify item level rater effects. For example, if three rater groups are tested, a significant fixed chi- square statistic means that at least one of these three rater groups is more severe/ lenient in their ratings on the overall scale. *Paired comparisons* are then performed to identify which rater group is significantly more severe / lenient in their ratings, or to show where (which items) the raters rate significantly more severe/lenient. We used an alpha level of 0.05 to determine a significant rater effect.

Objective 2. We validated the SDBM to results from on-road driving evaluation.

Details of design, participants, procedure, measurement and data collections are listed above under Objective 1(see pages 10-12).

Data Analysis

We used Predictive Analytics SoftWare (PASW)Statistics 20 (SPSS Inc, 2012) and WINSTEPS computer program (Version 3.74.0) (Linacre, 2012a) to perform the analyses.

Descriptive statistics

For the drivers, we conducted a descriptive analysis and included demographic, driving history, health-related characteristics, clinical tests and on-road test data. For F/C the descriptive analysis included their demographics, their history as a passenger, and how their lifestyle would be impacted if the driver reduced or stopped driving.

ROC curve analysis

Area under the curve or AUC is a measure of a test's discriminability including positive and negative results. In this study, we viewed an AUC between 0.7 and 0.9 as having an acceptable magnitude (Streiner & Cairney, 2007). Most important for the SDBM, to be used as a potential screening tool to accurately classify the drivers who fail the on-road test, we wanted sensitivity to be high (>0.70). Generally, we wanted to minimize misclassification of drivers, or false positives and false negatives. We generated the ROC curve and AUC estimates with PASW Statistics 20 (SPSS Inc, 2012) using measures. The measures were derived from the raw scores of the SDBM via Rasch analysis and are presented as logits (Bond & Fox, 2001; Classen et al., 2012a). Using the measure (logits) we present the ROC curves demonstrating five of these potential SDBM cut-point measures. Based on the cut-points we also calculated the associated specificity, sensitivity, error, PPV and NPV. The AUC of the ROC curve was based on a 95% confident interval (CI) and p-value ≤ 0.05 to indicate statistical significance.

Objective 3: Develop the instructional clinical outputs, or "keyforms"

Methods

Institutional review board approval was granted for this project. Participants provided written informed consent prior to focus group involvement and were paid $50 to $100 for participation.

Design

Stakeholder input was solicited via three focus groups, with each group addressing specific goals during different phases of developing a Web-based SDBM and keyform. The groups are listed below by stakeholder and purpose: *Focus Group 1*: Occupational therapy practitioners: To address keyform understandability and utility, and to obtain feedback on improving clarity. *Focus Group 2*: Expert Panel of CDRSs: To develop, from expert opinion, clinical recommendations for the caregiver/family members and to obtain feedback on the Web-based keyform. *Focus Group 3*: F/C: To obtain feedback on the understandability and ease-of-use of the Web-based SDBM and keyform.   ** Note: Focus Group 1 was conducted prior to the initiation of FDOT funding, however for continuity's sake it is being published in this report as those methods and findings supported the two focus groups conducted under FDOT funding.

Participants

We recruited participants by purposive sampling for all stakeholder groups (Morse, 1994). Sample size for the groups was set between 5 and 12 depending on the purpose and degree to which we required in-depth responses (Krueger, 2009). Specific criteria were: *Focus Group 1,* we recruited 12 occupational therapists via our networking with the AOTA Older Driver Group. Driving evaluators and occupational therapists with at least 2 years of clinical practice experience, who have completed driving screenings/ assessments/ evaluations, and who have worked with older adults ≥ 65 years. *Focus Group 2:* included an expert panel of five CDRSs, with at least 10 years of experience in driving evaluation and rehabilitation, and who have worked with older adults ≥ 65 years. *Focus Group 3:* included eight F/C, all previous participants who were selected based on gender, rural or urban residence, race (Caucasian, Asian or African-American), and relationship to the driver (spouse, adult child or friend).

Data Collection

We used set questions (focus group guide) to moderate the group and direct participant feedback. We asked participants about aspects of keyform utility, i.e., ease of use, time to complete, training required, format, interpretation, meaning, and relevance (Smart, 2006). Specific group content are next discussed.

Focus Group 1 (OTs): The setting was a private hotel conference room in Philadelphia, during the 2011 AOTA annual conference**.** The research team presented the development of the SDBM using the focus group guide and illustrated the keyform (see Figure 1). Moderated discussions were led by core research personnel, with participants divided into two groups. Designated research personnel took notes, and a representative from each group reported feedback which was audio-recorded and later transcribed.

Focus Group 2 (Expert Panel): The setting was at Adaptive Mobility Services, Orlando, FL. During the four hour expert panel meeting, members were oriented to the development and functionality of the keyform, and our goal to develop clinical recommendations. The use of the keyform was illustrated with three case study examples who had various outcomes following the on-road test (one fail, one was borderline, and one passed). The experts provided oral feedback, and completed 11 questions on keyform usability (overall ease of use of the keyform) via a visual analogue scale (VAS) (Streiner & Norman, 2008). We video recorded the panel discussion, for vivid retrieval of content during data analysis.

Focus Group 3 (F/Cs): The setting was a private conference room at the University of Florida, Gainesville, FL. The duration was approximately two hours and included an introduction to the Web-based SDBM and keyform, a discussion, and structured respondent feedback. Respondents had experience completing the proxy version SDBM in a previous study session. To orient them and introduce new developments, we described the Web-based SDBM and keyform, including the on-line scripts, instructions for administration and guidelines for interpretation. We presented a case study of a driver who failed, showed the recommendation and solicited feedback. We also discussed the recommendations for the two higher level drivers (“passed” and “borderline”). We created a focus group guide with questions targeted to F/C. In addition to answering the focus group guide questions, respondents were asked to suggest revisions and general feedback on the SDBM, keyform, and Web-based functions. Verbal feedback was audio-recorded for transcription and respondents provided written feedback on a VAS. Assigned research personnel took field notes, which were integrated with the verbal and written responses for data analysis.

Coding and Data Analysis

Focus Group 1 (OTs): We transcribed the focus group data and handwritten comments, verbatim, into Microsoft Word® documents and imported the documents into QSR International's NVivo 8 software (NVivo qualitative data analysis software, 2008) for coding. We used a directed content analysis approach (Hsieh & Shannon, 2005), a deductive approach to identify initial constructs for use as coding categories. We identified these initial constructs by, first, coding the questions posed to Focus Group members, and then coding the remaining data according to four broad themes that emerged from the data, i.e., *Face Validity*, *Appearance and Wording*, *Usability*, and *Recommendations for Improvement*. To ensure rigor, coding and results were reviewed by a primary and secondary analyst. Coding and results were reviewed in-depth, for rigor, by a primary and secondary analyst. Results were further refined by the research team. *Face Validity* refers to whether, in the respondent's judgment, the measure and items measure what they report to (e.g., driving behavior). Face validity for our study is the ability of the keyform or items to (1) discriminate between levels of driving ability, (2) highlight a driver's challenges, and (3) capture a driver's strengths and abilities. *Appearance and Wording* refers to the readability and visual appeal of the keyform (layout, font, spacing etc.), and whether or not the item language is clear. *Usability* refers to the overall ease-of-use of the keyform. Recommendations for *Improvement,* included suggestions for revisions, additions, and improved user friendliness for the OTs.

Focus Group 2 (Expert Panel): We transcribed the respondents' discussion and their handwritten comments. Using the directed content analysis approach (Hsieh & Shannon, 2005) we coded data to address the focus group discussions. From the data (VAS, video-taped materials and field notes) we synopsized changes to be made to the Web-based keyform (layout and descriptions), and we constructed texts for the clinical recommendations of the keyforms.

Focus Group 3 (F/C): We integrated the field notes, VAS responses, transcript and coded data to summarize responses using a directed content analysis approach (Hsieh & Shannon, 2005). From these responses, we identified recommendations to clarify wording, revise instructions, enhance usability of features (e.g., data entry via drop-down boxes rather than the type in method), improve the introductory script, and modify the presentation of the keyform. We also received feedback on the implications of the recommendations, such as need for follow-up conversations with the physician, or conflict arising between the driver and family based on ratings.

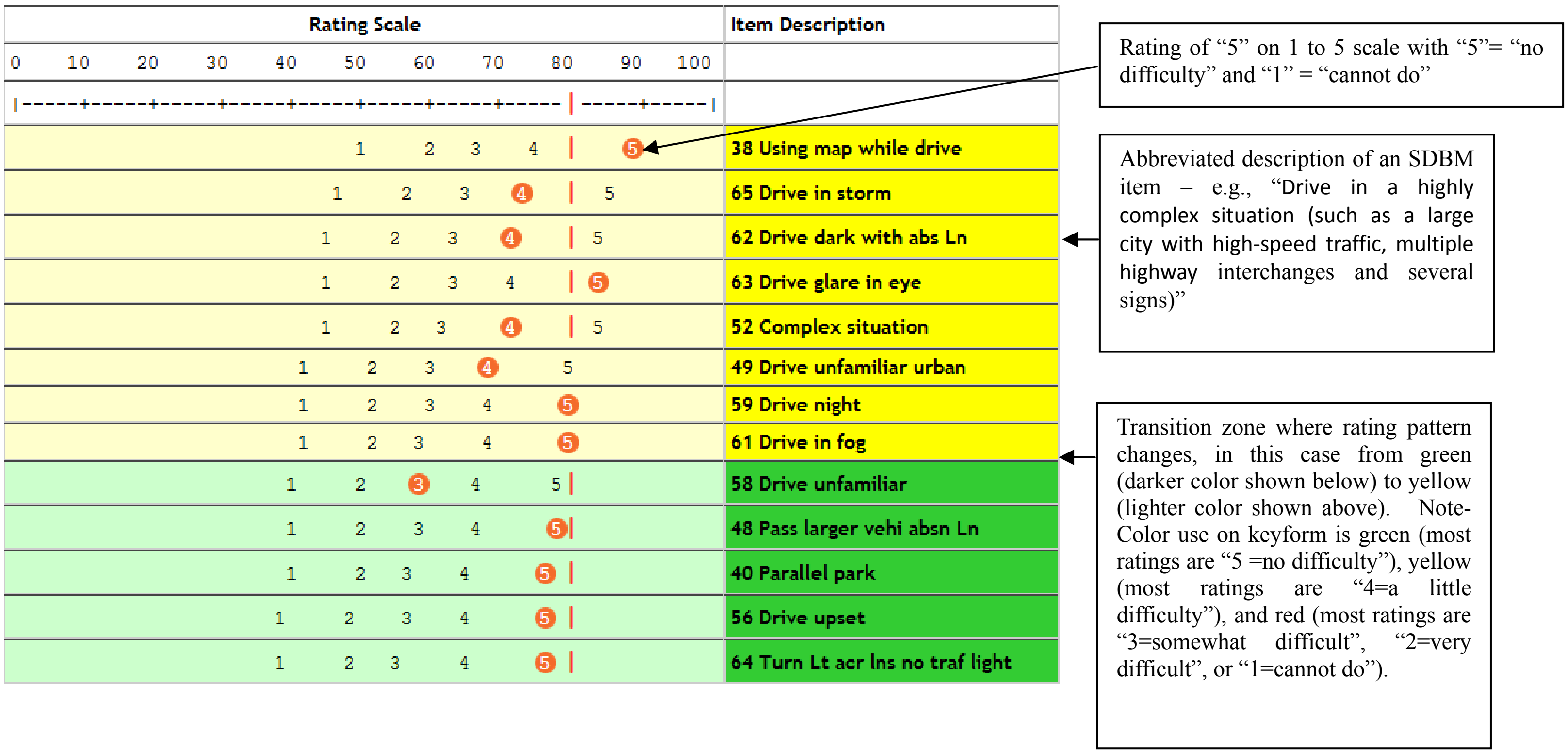


Figure 1. Example Keyform of a Driver Who "Passed" the On-road Test. Rating is by his/her family member/ caregiver. Ratings are mostly 4's (a little difficult) or 5's (no difficulty) with only one 3 (somewhat difficult). Note: In the Web-based version items are fully displayed when the cursor points to the items as listed in the "item description". Legend for abbreviations: vehi= vehicle; absn= absence; Ln= lane; Lt= left; acr= across; lns= lanes; traf= traffic

Objective 4. Develop a data collection system

This objective was achieved by using the steps below:

1. We built the Web-based SDBM and keyforms using PHP computer language (http://www.php.net/manual/en/index.php). The Web-based keyform replicates the WINSTEPS output Tables 2.2 & 2.12 (Linacre, 2012b). To better illustrate the difficulty ratings shown on the Web-based keyform, we developed an algorithm to color-code the form. We set default visual clustering values: green (most ratings are "5 = No difficulty"), yellow (4 = "A little difficult"), and red (most ratings are 3= "Somewhat difficult", 2="Very difficult", or 1= "Not difficult"). As shown in Figure 1, the algorithm identified a transition boundary as indicated by the color changes. The algorithm devises an average score for each item "row" by averaging 5 data points, the rating of two items above, selected item, and two items below. For example, a value of "5" represents the green zone, so when the value changes to "4" for a given row, a transition occurs and the color changes to yellow for that zone.

2. We used the item difficulties to calculate the person measure/ability, using the formula:

Formula to Calculate the Person Ability Estimate Using the Item Difficulties

$$b_r^{(t+1)} = b_r^{(t)} - \frac{r - \sum_i^L \sum_{k=1}^{m_i} k P_{rik}^{(t)}}{-\sum_i^L \left[\sum_{k=1}^{m_i} k^2 P_{rik}^{(t)} - \left(\sum_{k=1}^{m_i} k P_{rik}^{(t)} \right)^2 \right]}$$

, where r is the raw score

$$P_{nix} = \frac{exp \sum_{j=0}^{x} \left(B_n - D_{ij} \right)}{\sum_{k=0}^{m_i} exp \sum_{j=0}^{k} \left(B_n - D_{ij} \right)}$$

Pnix is the response probability of the user selecting the response x. Bn is the ability, and Dij is the item difficulty parameter/theta from WINSTEPS (Linacre, 2012b) we input for each questions. Bn is recalculated sequentially with each added item, starting from the easiest item to hardest item, and stops at the end.

Procedure for Setting the Cut-Scores

Based on measurement theory, psychometrics of the SDBM, exemplar cases, and team input, three driver classifications were established. Three driver classifications were basic driver, routine driver and accomplished driver as defined below, plus a category for each driver group who could not be classified (scoring did not fit the probability model) based on the ratings. However, each driver group(A-D) had two sub-classifications: one group who did not make critical driving (e.g., Group A1) errors, and the other who did make critical driving errors (e.g., Group A2). Two cut scores or thresholds were established using the F/C ratings for the group of 200 drivers. The first threshold separated the lowest rated drivers (basic drivers) from the moderately rated drivers (routine drivers). The second threshold separated the moderately rated drivers (routine drivers) from the highest rated drivers (accomplished drivers). In addition to the three categories based on difficulty ratings, it was necessary to establish a category (Group D- unable to categorize) for drivers whose rating pattern (based on Infit MnSq and Outfit MnSq) did not fit the Rasch model as the F/Cs' ratings showed unexpected or erratic patterns. The driver categories are as follows:

Group A 1 = Accomplished Driver: someone who is able to perform complex driving skills and may only experience difficulty with the most challenging skills.
Group A 2 = Accomplished Driver-Difficulty with one or more critical driving errors: someone who is able to perform complex driving skills, however, one or more critical driving errors were reported indicating a need for intervention.
Group B 1 = Routine Driver: someone showing difficulty with routine driving skills and early signs of needing intervention.
Group B 2 = Routine Driver: Difficulty with one or more critical driving errors: someone showing difficulty with routine driving skills and due to one or more critical driving errors there are critical safety concerns that need immediate attention.
Group C 1= Basic Driver: Although driver can still perform basic driving skills, there are safety concerns that need immediate attention.
Group C 2= Basic Driver: Difficulty with one or more critical driving errors: Although driver can still perform some basic driving skills, there are critical safety concerns that need immediate attention.
Group D1= Unable to classify: this driver's rating pattern could not be matched to a category.
Group D2= Unable to classify: this driver's rating pattern could not be matched to a category, due to one or more critical driving errors there are safety concerns that need immediate attention.

As part of setting the thresholds, we used three exemplar cases, choosing one for each category that was most representative based on F/C ratings and on-road test results (pass or fail). The research team looked at the rating patterns to determine the "transition zone boundary" where the F/C rating started changing from one category of difficulty to the next category (e.g., from rating of “No difficulty” to rating of “A little difficulty”) as these items provided information for setting category thresholds. Items were calibrated based on Rasch analyses and charted to demonstrate the item hierarchy (easy to difficult) and the pattern of difficulty ratings (from not difficult to very difficult). Using the functional stages method (Jette et al., 2007), we identified difficulty patterns associated with each driver category. For example, a basic driver may be rated “not difficult” for the easiest (basic) driving skills, but rated as “somewhat difficult” or “very difficult” for routine driving skills, indicating a high level of overall driving difficulty. In contrast, the routine driver is rated “not difficult” for the basic driving, but ratings for the routine driving items indicated moderate difficulty, and ratings for the accomplished (hardest) items indicated moderate to severe difficulty. Lastly, the accomplished drivers are rated “not difficult” for both the basic and routine driving items, but may be rated as having slight to moderate difficulty with the most challenging driving skills.

To further identify and classify at-risk drivers across the groups, the team identified seven critical driving errors (see Figure 2). If a F/C gave a rating of “somewhat difficult” or “very difficult” for one or more critical items, the driver was labeled as having made a critical driving error. The SDBM output included a driver profile with keyform to illustrate ratings, driver category, examples of difficult items, and category specific safety recommendations. Drivers in Group C and drivers with critical errors received the strongest clinical recommendation to pursue a CDE. The recommendations for the three main groups are:

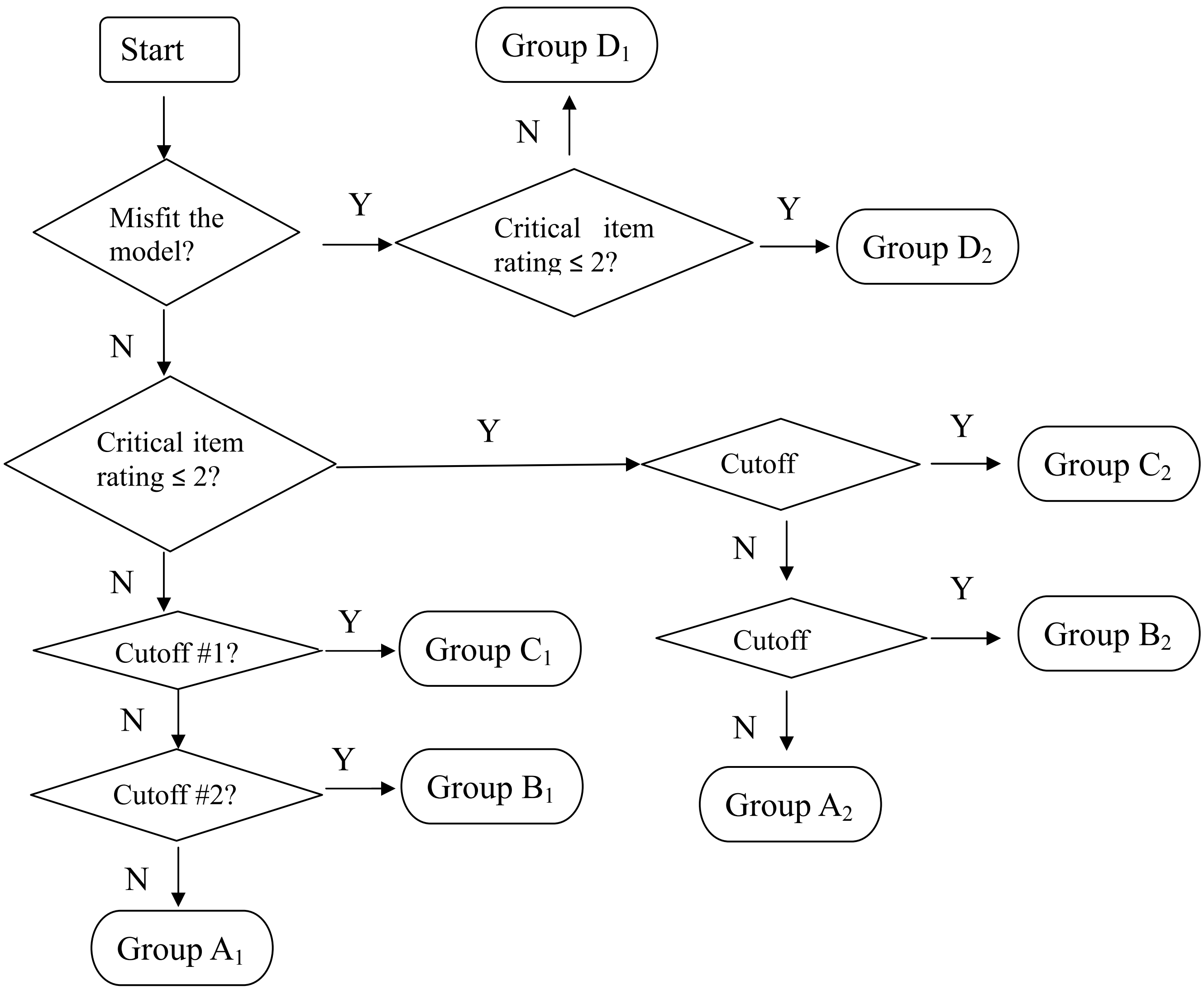
Basic Driver: **<u>Specific Recommendations</u>**: We recommend the driver see a doctor for a physical exam as soon as possible and not drive until he/she is able to undergo a comprehensive driving evaluation conducted by a Certified Driver Rehabilitation Specialist. Information on the use and access to alternative transportation (other than the personal automobile) may be available from the local Area Agency on Aging. **<u>General Recommendations</u>**: Based on guidelines of The American Geriatrics Society, we recommend an eye exam annually or earlier if there are changes in health or vision.

Routine Driver: **<u>Specific Recommendations</u>**: We recommend a doctor's appointment to start a conversation about conditions that may impact driving safety. The driver will also benefit from a comprehensive driving evaluation to address safety concerns. We recommend repeating this self-screening annually or when the driver experience changes in health or functional status. **<u>General Recommendations</u>**: The American Geriatrics Society recommends a physical and eye exam annually or earlier when needed. We recommend taking a class for mature drivers, such as those offered by AAA, AARP, or a local driving school.

Accomplished Driver: **<u>Specific Recommendations</u>**: It may be helpful to avoid or limit driving situations that are challenging. Based on your ratings, we do not think that a comprehensive driving evaluation is critical at this time. We recommend repeating this self-screening annually or when the driver experiences changes in health or functional status. **<u>General Recommendations</u>**:

Additionally, The American Geriatrics Society recommends a physical and eye exam annually or earlier when needed. We recommend taking a class for mature drivers, such as those offered by AAA, AARP, or a local driving school.

Figure 2. Flowchart for Driver Categorization based on F/Cs' Ratings



- Group D1 – misfit and no critical items
- Group D2 – misfit with critical items
- Group C1 – lowest ability pattern
- Group C2 – lowest ability pattern with critical items
- Group B1 – middle ability pattern
- Group B2 – middle ability pattern with critical items
- Group A1 – highest ability pattern
- Group A2 – highest ability pattern with critical items

Critical items

- 9 Stay in Ln
- 27 Change Ln mod traf
- 30 Maintain Ln when turn
- 32 Turn Rt enter traf
- 39 Let turn into traf
- 41 Stay within Ln mark
- 42 Stay within Ln absn

CHAPTER 4 - RESULTS

Demographics of the Participants

Table 1 presents the key demographics of 200 drivers and 200 F/Cs, applicable for Objectives 1, 2, and 4. The demographics of the focus group respondents for Objective 3 are described separately.

Table 1. Demographics and Driving Characteristics of Older Drivers and their F/C

	Older driver (N = 200)	**Family members/caregivers** (N = 200)
Age mean *(SD)* year	72.64 (5.35)	62.44 (14.76)
Age range year	65-85	18-85
Gender: Male	110 (55.0%)	55 (27.5%)
Race		
Caucasian	177 (88.5%)	180 (90.0%)
African-American	12 (6%)	12 (6.0%)
Others	11 (5.5%)	8 (4%)
Education		
College or university	114 (57.0%)	93 (46.5%)
Vocational /Associate Degree	43 (21.5%)	75 (37.5%)
≤ High school	43 (21.5%)	32 (16.0%)
Drive 7 days/week	102 (51.0%)	27 (13.5%)[a]
Licensed driver	200 (100%)	197 (98.5%)
Living alone	52 (26.0%)	25 (12.5%)
Living with partner/spouse	129 (64.5%)	111 (55.5%)[b]
MMSE mean *(SD)*	27.99 (1.84)	NA
Self reported number of medications used	6.73 (4.45)	NA

Note: MMSE = Mini Mental State Examination; NA= Not applicable; SD = standard deviation
Median of age for F/C = 67.0;
[a]: Ride with the driver 7 days per week.
[b]: The relationship with the driver is spouse or partner.

Objective 1. Determine the SDBM measurement properties

Task 1(a): Confirm the factor structure of the SDBM.

We conducted confirmatory factor analysis (CFA) to examine the unidimensionality assumption of Rasch analysis using Mplus 5.1 (Muthen & Muthen, 2006). We found that 54 items of the SDBM assessed one overarching construct. We have used 3 criteria to evaluate the unidimensionality: comparative fit indices (CFI) ≥ 0.90; Tucker-Lewis Indices (TLI) ≥ 0.90; and root mean square error of approximations (RMSEA) < 0.08 (Brown, 2006), with results shown in Table 2.

Per CFA criteria, we excluded 14 items (item numbers 1-8, 10, 11, 13, 14, 16, 17) as some items did have a ceiling effect with ≥ 95% of the ratings on the items representing a "5". Due to the skewed ratings, these items did not contribute to the safe driving construct. As indicated in Table 2, the values of CFI and TLI ranged from 0.86 to 0.98. F/Cs and Evaluator groups' ratings fit a one-factor model based on both the CFI and TLI. However, the values of RMSEA did not represent a good model fit. Overall, most of the values indicated our data fit the one-factor model so we concluded that 54 items of the SDBM met the unidimensionality assumption. From the CFA table, the caregiver fit the one-factor model the best, followed by the evaluator, and the driver having the least satisfactory fit.

Table 2. Factor Analysis for Three Rater Groups

	Driver	**Family member/caregiver**	**Evaluator**
Chi-square	292.89	179.00	401.20
Df	60	60	50
p	0.000	0.000	0.000
CFI	0.862	0.942	0.926
TLI	0.910	0.979	0.952
RMSEA	0.139	0.10	0.187

Note: F/C = Family member/caregiver, Df = Degree of freedom, *p* = *p*-value, CFI = Comparative Fit index, TLI = Tucker-Lewis index, RMSEA = Root Mean Square Error of Approximation.

Task 1(b): Describe the item and person level properties of the SDBM.

Rasch analysis rating scale structure

After we have collapsed the rating scales, the results showed that the Outfit MnSq for category 1 "Very difficult" for the evaluator group was 2.12 logits. Using the criterion, Outfit MnSq < 2.0), we surmised that the evaluators rated in an unexpected fashion pertaining to category 1. As expected, the average measure of each rating category increased in logits e.g., the average measure of category 4 was higher than that of category 3. This finding illustrated that the complexity from one rating category (e.g., evaluator group rating category 2 = -0.11) increased in difficulty to the next category (e.g., evaluator group rating category 3 = 1.70, and in category 4 = 4.82).

Item and Person Statistics

We ran Rasch analyses on the SDBM for each of the three groups: drivers, F/Cs, and evaluators (Table 3).

The *item statistics* of Rasch analysis showed 3 - 10 (5.6% - 18.5%) misfitting items across three groups of raters when using criterion A (Infit MnSqs were from 0.6 to 1.4, and the previous standardized fit statistics from -2 to 2), and 1 – 5 (1.9 – 9.3%) when using criterion B (Infit MnSqs were from 0.5 to 1.7 and the standardized fit statistics from -2 to 2). The evaluator group had the highest number (18.52 % / 9.26%) of misfitting items, and the driver group having the lowest number of misfitting items. One item, # 38 “Use a map while driving” showed high infit statistics (misfit), or unexpected rating patterns, for both driver and F/C groups. However, # 38 did not misfit the ratings of the evaluator group. Instead, items #19, 20, 35, 43, and 44 showed high infit statistics on the ratings of the evaluator group, with item #44 “Look left and right before crossing an intersection” had the highest value. This indicated that the item was mis-performing as rated by the evaluator group. Table 2 also illustrates that all three groups showed good item reliability (0.97 – 0.99) and good item separation (5.43 – 8.45). Therefore, the item statistics were good, except for the Evaluator item fit (Criterion A).

The *person statistics* of Rasch analysis showed 13 - 24 (6.5% - 12%) misfitting persons across the three groups of raters when using criterion A (Infit MnSqs were from 0.6 to 1.4, and the standardized fit statistics from -2 to 2), and 8 – 13 (4% – 6.5%) when using criterion B (Infit MnSqs were from 0.5 to 1.7 and the standardized fit statistics from -2 to 2). The evaluator group had the highest number of person misfit (A=12.5%; B=6.5%), illustrating that the evaluator group provided a different response pattern compared to the other two groups. The results showed good person reliability (0.90 – 0.96), and good person separation (3.07 – 4.68) across the three rater groups. Person means in logits, defined as the average of the older drivers’ abilities rated by three rater groups, were about two standard deviations higher than the item means across three groups. This finding indicates that the average abilities of the drivers exceeded the average difficulty of the items. Ceiling effects were evident for the older driver (6.5%) and F/C groups (14.0%). Overall person statistics were good, except for the evaluator group.

Table 3. Rasch Analysis Summary - 54-item SDBM for the Driver, F/C, and Evaluator

	Driver	**Family member/ caregiver**	**Evaluator**
Item Misfitting (A)	3 (5.56%)	2 (3.70%)	10 (18.52%)
Item Misfitting (B)	1 (1.85%)	1 (1.85%)	5 (9.26%)
Item with minimum estimate value	0	0	0
Item Reliability	0.98	0.97	0.99
Item Separation	6.5	5.43	8.45
Person Misfitting (A)	14 (7%)	13 (6.5%%)	24 (12%)
Person Misfitting (B)	8 (4%)	9 (4.5%)	13 (6.5%)
Person Reliability	0.92	0.90	0.96
Person Separation	3.43	3.07	4.68
Person Strata	4.91	4.43	6.57
Cronbach's Alpha	0.96	0.98	0.97
Person Mean (logits)	3.44	3.47	3.39
Standard Deviation (logits)	1.55	1.59	1.71
Ceiling/Floor	13 (6.5%) / 0	28 (14%) / 0	0 / 0

Note:
Misfit criterion A: INFIT MNSQ > 1.4 & INFIT Z-Score >2 || INFIT MNSQ < 0.6 & INFIT Z-Score <2
Misfit criterion B: INFIT MNSQ > 1.7 & INFIT Z-Score >2 || INFIT MNSQ < 0.5 & INFIT Z-Score <2
Strata = (4*Sep.+1)/ 3

Item Hierarchy

We are presenting the item map as rated by three rater groups in figures 3 through 5. This figure illustrated the drivers' abilities, as rated by the individual rater groups and item difficulties, on one linear continuum. On the left side of the continuum, "M", "S", and "T" indicated the mean, one standard deviation (SD) and two SD from the mean of the older drivers' abilities. On the right side of the continuum, "M", "S", and "T" indicated the mean, one SD and two SD from the mean of the item difficulties. In this study, we anchored the mean difficulty of 54 items to zero. The item map for the F/C group (figure 4) showed that the easiest item was item # 44 " Look left and right before crossing an intersection", and the most challenging items were # 68 "Drive on an icy road" and # 38 "Use a map while driving". Item # 68 and # 38 were also the two most challenging items for the driver and evaluator groups, while the easiest item for the driver group was item # 15 "Use the car controls", and for the evaluator group was item # 23 "Drive in light rain". The driver (Figure 3) and F/C (figure 4) item maps generally follow the same pattern as what we expected, but not the evaluator group. For example, item # 9 located as easy item on the driver and F/C item maps, but as medium difficult on the evaluator item map; item #29 located as medium easy item on the driver and F/C item maps, but as difficult on the evaluator item map.

In general, the item maps for the three groups of raters showed that the average of the drivers' abilities was more than two standard deviations higher than the average of items' difficulties. The distribution of the older drivers (left side) and items (right side) indicated that this sample had a relative high ability in terms of fitness-to-drive.

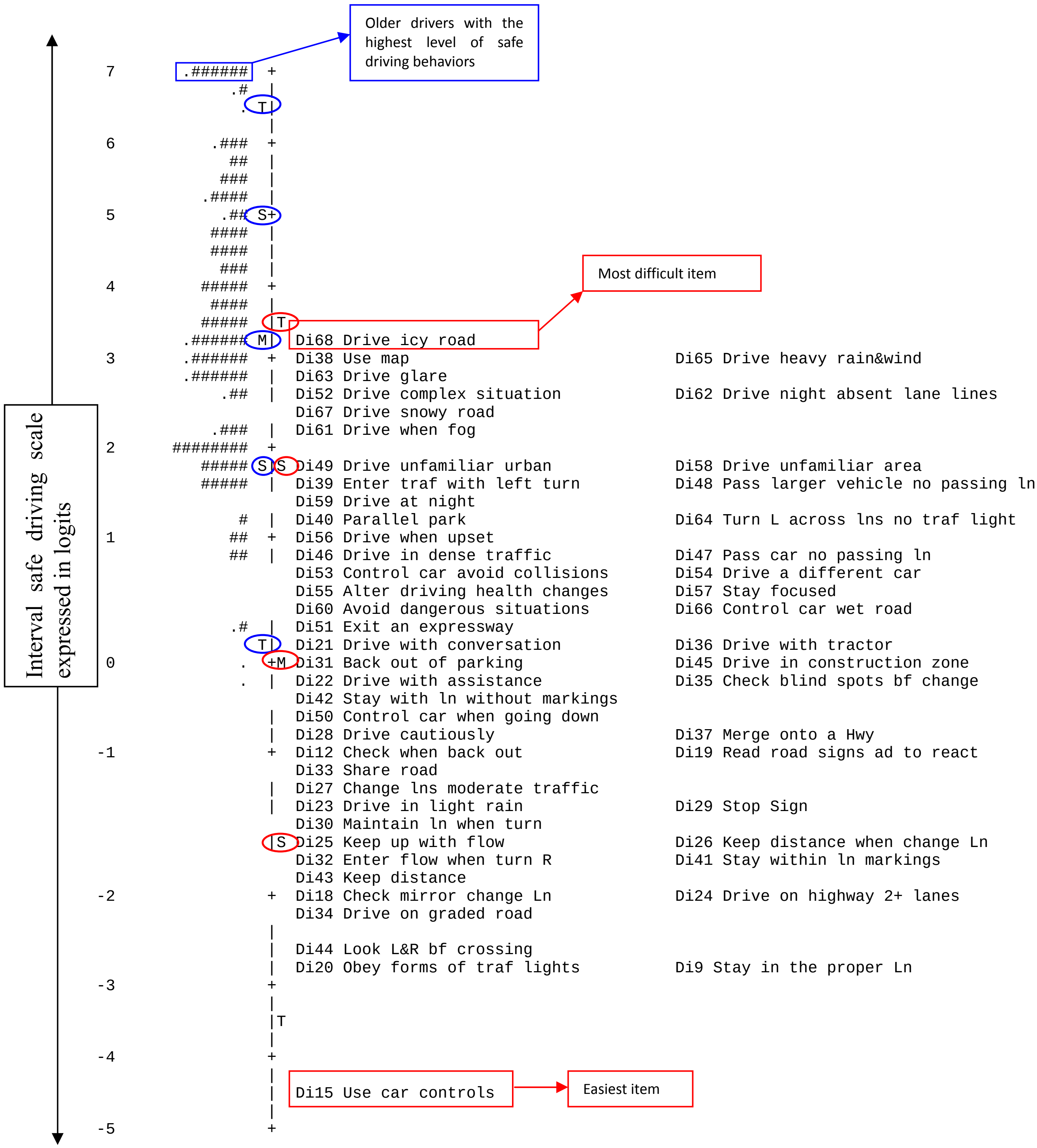


Figure 3: Item Map Illustrating Older Driver Group (Self-ratings) vs. Item Difficulty.

Note: Each '#' represents 2 drivers, and each "." represent 1 driver, M =Mean of person or item distribution, S = One standard deviation from the person or item mean, T =Two standard deviations from the person or item mean

Abbreviations: Di: Driver group item; Rd: road; Ln: lane; vehi: vehicle; absn: absence; acr: across; R or Rt: right; Pedestn: pedestrian; traf: traffic; convers: conversation; fr: from; L or Lt: left; bf: before

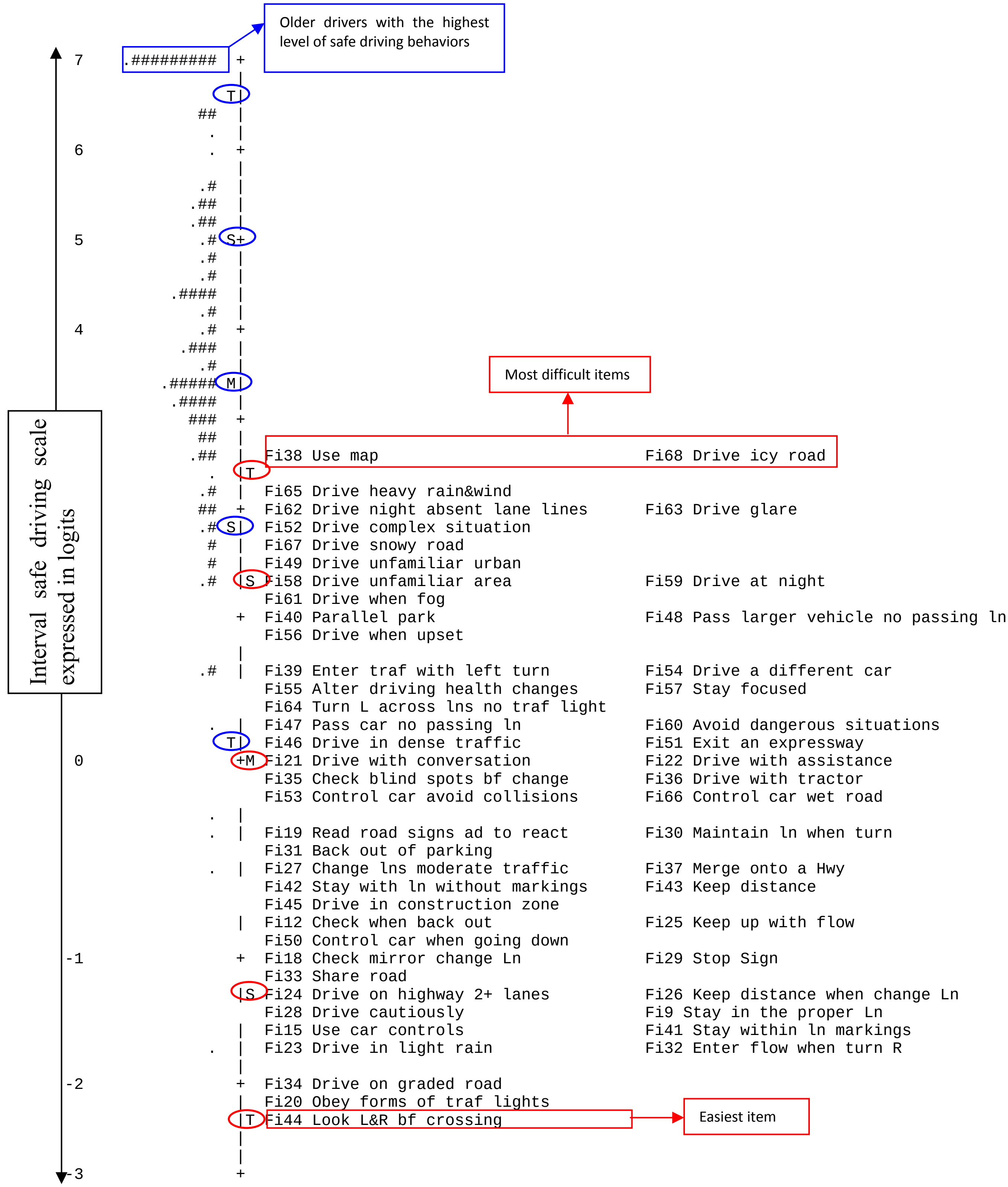


Figure 4: Item Map Illustrating Family Member/Caregiver Ratings vs. Item Difficulty.

Note: Each '#' represents 3 drivers, and each "." represent 1 driver, M =Mean of person or item distribution, S = One standard deviation from the person or item mean, T =Two standard deviations from the person or item mean.

Abbreviations. Fi: family member/caregiver item; Rd: road; Ln: lane; vehi: vehicle; absn: absence; acr: across; Rt: right; Pedestn: pedestrian; raf: traffic; convers: conversation; fr: from; Lt: left; bf: before

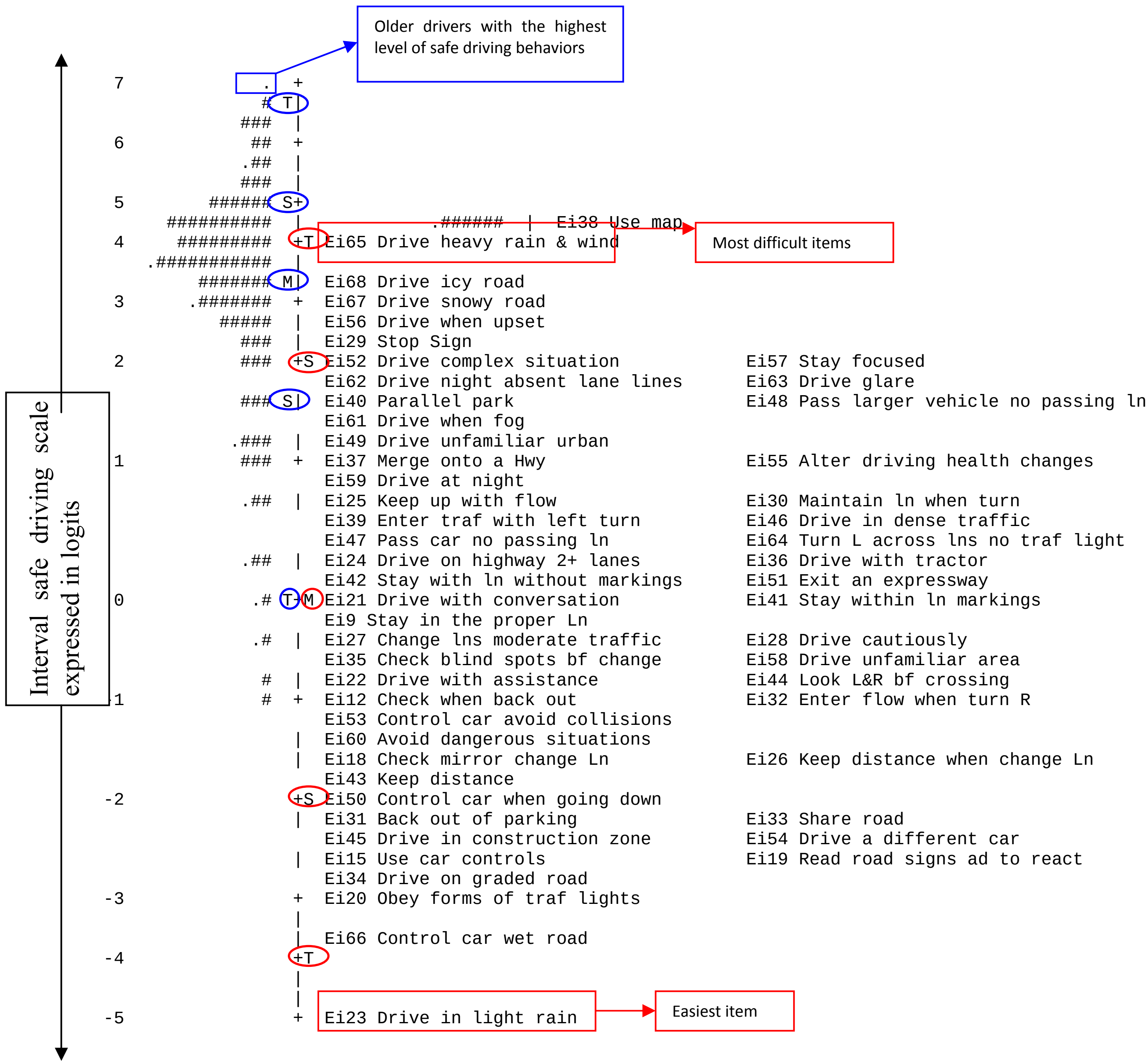


Figure 5. Item Map Illustrating Evaluator Ratings vs. Item Difficulty.

<u>Note:</u> Each '#' represents 2 drivers, and each "." represents 1 driver, M =Mean of person or item distribution, S = One standard deviation from the person or item mean, T =Two standard deviations from the person or item mean

<u>Abbreviations</u>. Ei: evaluator item; Rd: road; Ln: lane; vehi: vehicle; absn: absence; acr: across; Rt: right; Pedestn: pedestrian; traf: traffic; convers: conversation; fr: from; Lt: left; bf: before

Task 1(c): Determine the rater severity of the three rater groups.

Inter-rater reliability

The ICC among the ratings of three rater groups was significant but weak (ICC = 0.253, p<0.001, 95% CI = 0.164, 0.345). The significant correlation on the 41 items was observed between the ratings of the evaluator and the F/C groups (ICC = 0.394, p < 0.001, 95% CI = 0.270, 0.505), between the ratings of the older driver and the F/C groups (ICC = 0.141, p = 0.023, 95% CI = 0.002, 0.247), and between the older driver and the evaluator groups (ICC=0.169, p = 0.008, 95% CI = 0.031, 0.301).

Rater Effects

Facet ruler of the SDBM

Figure 5 depicts three Facets (raters, drivers, items) on the linear interval scale for the SDBM. Based on Facets procedures, 33 items of the SDBM were used to establish the ruler while 35 items did not meet the rating criteria due to variability of responses for this analysis. The first column, titled measure, is the interval scale expressed as a logit unit. The second column displays the severity of raters from lenient (bottom) to severe (top). The third column shows the distribution of the safe driving ability of the drivers, from bottom to top representing the drivers with poor safe driving abilities to good driving abilities. The fourth column displays item difficulties, from bottom to top representing that the items were essentially easy and then progress to levels of increasing difficulty. The fifth column shows the likelihood of applying the rating scale in relation to the raters' abilities; that is, when a driver's estimated ability is between 5 and 6 logits, he/she will likely receive a rating of 4 on this measure. In the second column, the driving evaluator is located above the caregiver, indicating that driving evaluator is the more severe rater. The distribution of the drivers' abilities was on the upper part of the ruler as displayed in the third column while the distribution of the item difficulties were on the lower part of the ruler as displayed in the fourth column. This indicated that the drivers had, generally speaking, high levels of safe driving abilities, as measured with this 33-Item scale.

```
+-----------------------------------------------------------------------------------------------------------+
|Measr|-Raters|+Persons                                              |-Items                           |SDBM |
|-----+-------+-----------------------------------------------------+---------------------------------+-----|
|   6 +       + 81                                                  +                                 + (4) |
|     |       |                                                     |                                 |     |
|     |       | 60  94  132                                         |                                 |     |
|     |       |                                                     |                                 |     |
|     |       |                                                     |                                 |     |
|     |       |                                                     |                                 |     |
|     |       |                                                     |                                 |     |
|   5 +       + 41  76  80  83  89  96  100                         +                                 +     |
|     |       |                                                     |                                 |     |
|     |       | 74  95  104 111 172                                 |                                 |     |
|     |       | 43                                                  |                                 |     |
|     |       | 17  50  55  99  121                                 |                                 |     |
|     |       |                                                     |                                 |     |
|     |       | 44  59  160 177 194                                 |                                 |     |
|   4 +       + 34  37  63  90  101 136 171                         +                                 +     |
|     |       | 107 108 151                                         |                                 |     |
|     |       | 79  113 115 134 195                                 |                                 |     |
|     |       | 56  72  159                                         |                                 |     |
|     |       | 39  77  117 167 178 180                             |                                 |     |
|     |       | 29  47  92  143                                     |                                 |     |
|     |       | 64  66  125                                         |                                 |     |
|   3 +       + 10  22  26  40  53  58  103 131 148 152 169 173 179 +                                 +     |
:     :       : 199                                                 :                                 :     :
|     |       | 11  48  61  84  91  110 122 127 133 147 154 157 181 |                                 |     |
:     :       : 182 188 197                                         :                                 :     :
|     |       | 87  93  106 116 141 158 170 174                     |                                 |     |
|     |       | 52  65  67  68  86  130 135 156                     |                                 |     |
|     |       | 4   24  27  49  62  69  88  98  112 120 140 146 155 |                                 |     |
:     :       : 162                                                 :                                 :     :
|     |       | 20  23  38  114 126 164 192                         |                                 |     |
|     |       | 1   12  35  70  82  97  124 145 163 166 196         |                                 |     |
|   2 +       + 54  73  75  78  129 139 161 168 176 184 185 186 200 +                                 +     |
|     |       | 13  21  25  109 128 137 165 183 190 198             |                                 |     |
|     |       | 6   14  33  57  118 123 144                         | Drive complex situation         |     |
|     |       | 15  30  31  32  42  51  71  85  102 105 119 142     |                                 |     |
|     |       | 16  36  149 191                                     |                                 | --- |
|     |       | 28  46  138 153                                     | Drive unfamiliar urban          |     |
|     |       | 8                                                   | Stay focused                    |     |
|   1 +       + 5   7   19  187                                     +                                 +     |
|     |       | 2   3   18  45  150 189                             | Drive unfamiliar area           |     |
:     :       :                                                     : Enter traf with left turn       :     :
:     :       :                                                     : Turn L across lns no traf light :     :
|     |       | 175                                                 | Pass car no passing ln          |     |
|     |       | 193                                                 | Drive in dense traffic          |  3  |
:     :       :                                                     : Stop Sign                       :     :
|     | E     |                                                     | Drive with tractor              |     |
|     |       | 9                                                   | Avoid dangerous situations      |     |
:     :       :                                                     : Drive with conversation         :     :
:     :       :                                                     : Merge onto a Hwy                :     :
:     :       :                                                     : Stay with ln without markings   :     :
|     |       |                                                     | Keep up with flow               |     |
:     :       :                                                     : Maintain ln when turn           :     :
*   0 *       *                                                     * Check blind spots bf change     * --- *
:     :       :                                                     : Drive with assistance           :     :
|     | D     |                                                     | Change lns moderate traffic     |     |
:     :       :                                                     : Drive on highway 2+ lanes       :     :
|     | FC    |                                                     | Back out of parking             |     |
:     :       :                                                     : Stay in the proper Ln           :     :
:     :       :                                                     : Stay within ln markings         :     :
|     |       |                                                     | Check when back out             |     |
|     |       |                                                     | Read road signs ad to react     |  2  |
|     |       |                                                     | Enter flow when turn R          |     |
:     :       :                                                     : Keep distance                   :     :
:     :       :                                                     : Look L&R bf crossing            :     :
|     |       |                                                     | Check mirror change Ln          |     |
:     :       :                                                     : Keep distance when change Ln    :     :
:     :       :                                                     : Share road                      :     :
|  -1 +       +                                                     +                                 +     |
|     |       |                                                     |                                 |     |
|     |       |                                                     |                                 |     |
|     |       |                                                     |                                 | --- |
|     |       |                                                     | Use car controls                |     |
|     |       |                                                     | Obey forms of traf lights       |     |
|     |       |                                                     |                                 | (1) |
|  -2 +       +                                                     +                                 +     |
|-----+-------+-----------------------------------------------------+---------------------------------+-----|
|Measr|-Raters|+Persons                                              |-Items                           |Scale|
+-----------------------------------------------------------------------------------------------------------+
```

Figure 6. Facet Ruler of the 33-item SDBM

Legend: Measr= measure; FC= family member/caregiver; E=driving evaluator; Each number represents a person.

Fit statistics of the rater groups. The infit MnSqs and the outfit MnSqs for both rater groups were between 0.93 and 1.15, well within the defined criteria of 0.6 and 1.4 (Bond & Fox, 2001; Linacre, 2002a; Wright & Linacre, 1994).

The fixed chi-square. The *fixed chi-square* value, 586.1with 2 degree of freedom, was statistically significant ($p<0.001$). The overall ratings among three rater groups showed significant rater effects.

Paired comparisons. The results of the *paired comparisons* showed significant rater effects on 19 items (Table 4 and Figure 6). Although the ratings of the evaluators were more severe on the overall scale, the F/C group rated 9/19 items more severely when compared to the evaluator group.

Table 4. Items with Significant Rater Effects

F/Cs are more severe raters than Evaluators on nine items:	***Measure F/C***	***Measure Evaluator***	***Contrast***	***Joint S.E.***	***T***	***P-value***
15. Use car controls	-0.69	-1.66	0.97	0.36	2.73	0.0065
19. Read road signs ad to react	0.13	-1.6	1.73	0.30	5.67	< 0.0001
22. Drive with assistance	0.34	-0.19	0.53	0.21	2.55	0.0111
31. Back out of parking	0.13	-1.36	1.5	0.28	5.28	< 0.0001
33. Share road	-0.43	-1.42	0.98	0.32	3.1	0.0021
35. Check blind spots bf change	0.31	-0.14	0.45	0.21	2.17	0.0305
43. Keep distance	-0.08	-0.9	0.82	0.26	3.14	0.0018
58. Drive unfamiliar area	1.22	0.04	1.18	0.17	6.92	< 0.0001
60. Avoid dangerous situations	0.63	-0.68	1.31	0.22	6.02	< 0.0001
Evaluators are more severe raters than F/Cs on ten items:	**Measure F/C**	**Measure Evaluator**	**Contrast**	**Joint S.E.**	**T**	**P-value**
9. Stay in the proper Ln	-0.53	0.24	-0.76	0.25	-3.02	0.0027
24. Drive on highway 2+ lanes	-0.48	0.41	-0.89	0.25	-3.62	0.0003
25. Keep up with flow	-0.19	0.61	-0.79	0.22	-3.61	0.0003
29. Stop Sign	-0.39	1.45	-1.84	0.23	-8.07	< 0.0001
30. Maintain ln when turn	0.07	0.50	-0.42	0.20	-2.09	0.0375
37. Merge onto a Hwy	-0.08	0.77	-0.85	0.21	-4.05	0.0001
41. Stay within ln markings	-0.75	0.13	-0.88	0.28	-3.19	0.0015
42. Stay with ln without markings	-0.08	0.41	-0.49	0.22	-2.3	0.0222
44. Look L&R bf crossing	-1.40	-0.27	-1.13	0.37	-3.09	0.0022
57. Stay focused	0.77	1.42	-0.65	0.16	-4.09	0.0001

Legend: SE= Standard Error, F/C = family member/caregiver, ad= advance, bf = before, Hwy = highway, ln = lane, L = left, R = right.

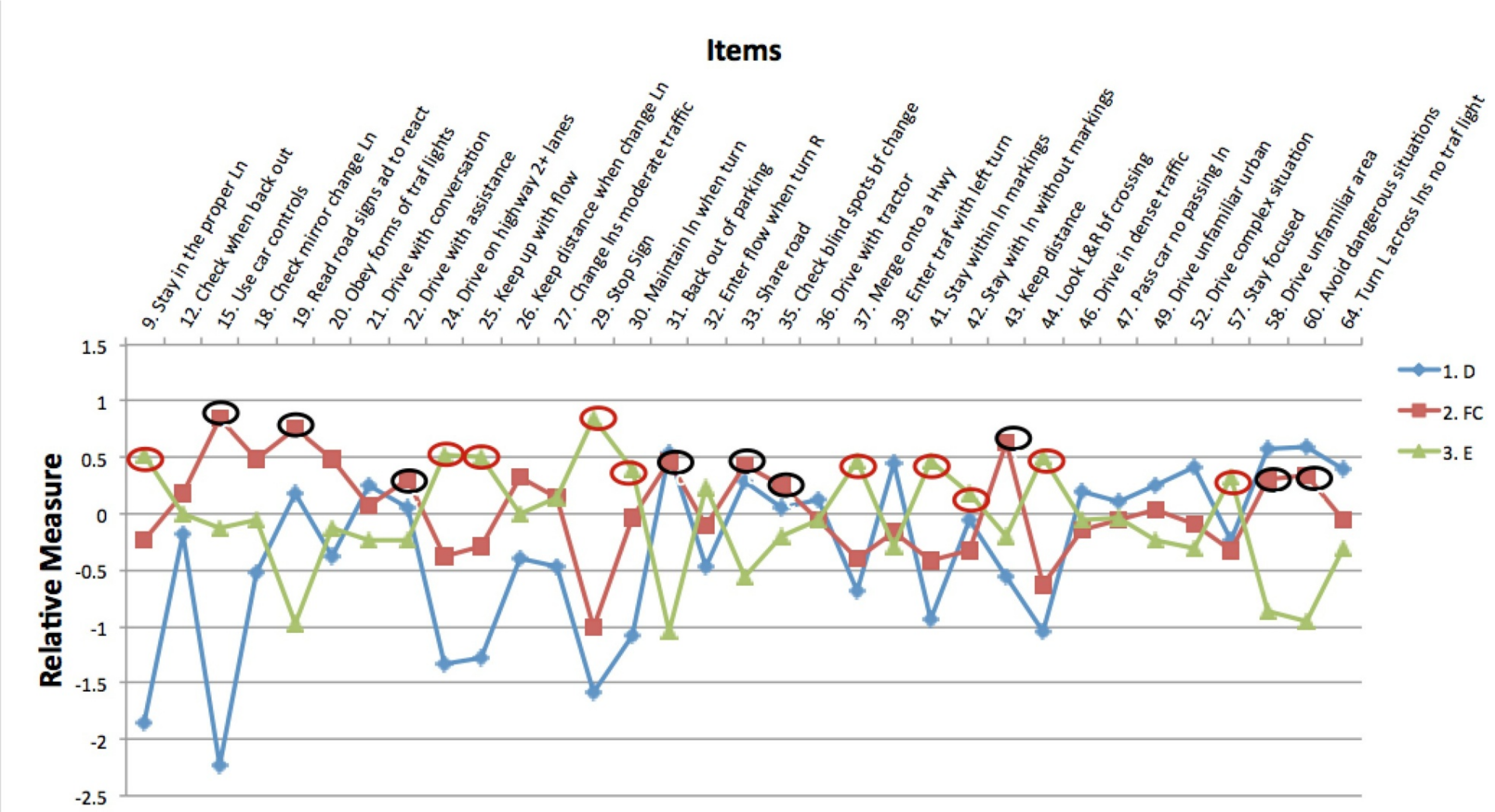


Figure 7. Bias Analysis Map for the Evaluator and Family Member/Caregiver Rater Groups (Item N = 33, Rater N = 200)

Legend: C=Family member/caregiver; E=Driving evaluator. The map shows significant rater effects on 19 items: #9, #15, #19, #22, #24, #25, #29, #30, #31, #33, #35, #37, #41, #42, #43, #44, #57, #58, #60. The ratings of the F/C group were significantly more severe than the ratings of the evaluator group on 9 items: #15, #19, #22, #31, #33, #35, #43, #58, #60, while the ratings of the evaluator group were significantly more severe than the F/C group on item 7 item: #9, #24, #25, #29, #30, #37, #41, #42, #44, #57.

Objective 2: We validated the SDBM measure to results from on-road driving evaluation. Table 5 and Figure 7 showed the ROC curve and the AUC based on ratings of three rater groups. The AUC based on drivers' ratings is 0.617, 95% CI = (0.513, 0.720), $p = 0.039$. The AUC based on F/Cs' ratings is 0.726, 95% CI = (0.628, 0.825), $p < 0.001$. Table 6 (Driver ratings) and Table 7 (F/C ratings) provide the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for a series of five potential cut-points based on Logit scores.

Table 5. Area under the Curve for Three Rater Groups

	Area	*Standard Error*	*p -value*	*95% Confidence Interval*	
Driver	0.617	0.053	0.039	0.513	0.720
F/C	0.726	0.050	>0.001	0.628	0.825
Evaluator	0.978	0.008	>0.001	0.962	0.995

Table 6. Sensitivity, Specificity, PPV, and NPV based on Driver Ratings

Cutoff	*1=65.205*	*2=67.485*	*3=69.115*	*4=72.585*	*5=74.865*
Sensitivity	0.387	0.516	0.613	0.774	0.806
Specificity	0.716	0.645	0.568	0.473	0.402
PPV	0.200	0.211	0.207	0.212	0.198
NPV	0.864	0.879	0.889	0.920	0.919
Error	0.897	0.839	0.819	0.752	0.791
False Positive	48	60	73	89	101
False Negative	19	15	12	7	6
Total Misclassification	67	75	85	96	107

Table 7. Sensitivity, Specificity, PPV, and NPV based on F/C Ratings

Cutoff	*1=52.630*	*2=68.795*	*3=70.795*	*4=71.915*	*5=73.465*
Sensitivity	0.194	0.581	0.677	0.742	0.806
Specificity	0.982	0.763	0.680	0.633	0.604
PPV	0.667	0.310	0.280	0.271	0.272
NPV	0.869	0.908	0.920	0.930	0.944
Error	0.824	0.656	0.642	0.625	0.590
False Positive	3	40	54	62	67
False Negative	25	13	10	8	6
Total Misclassification	28	53	64	70	73

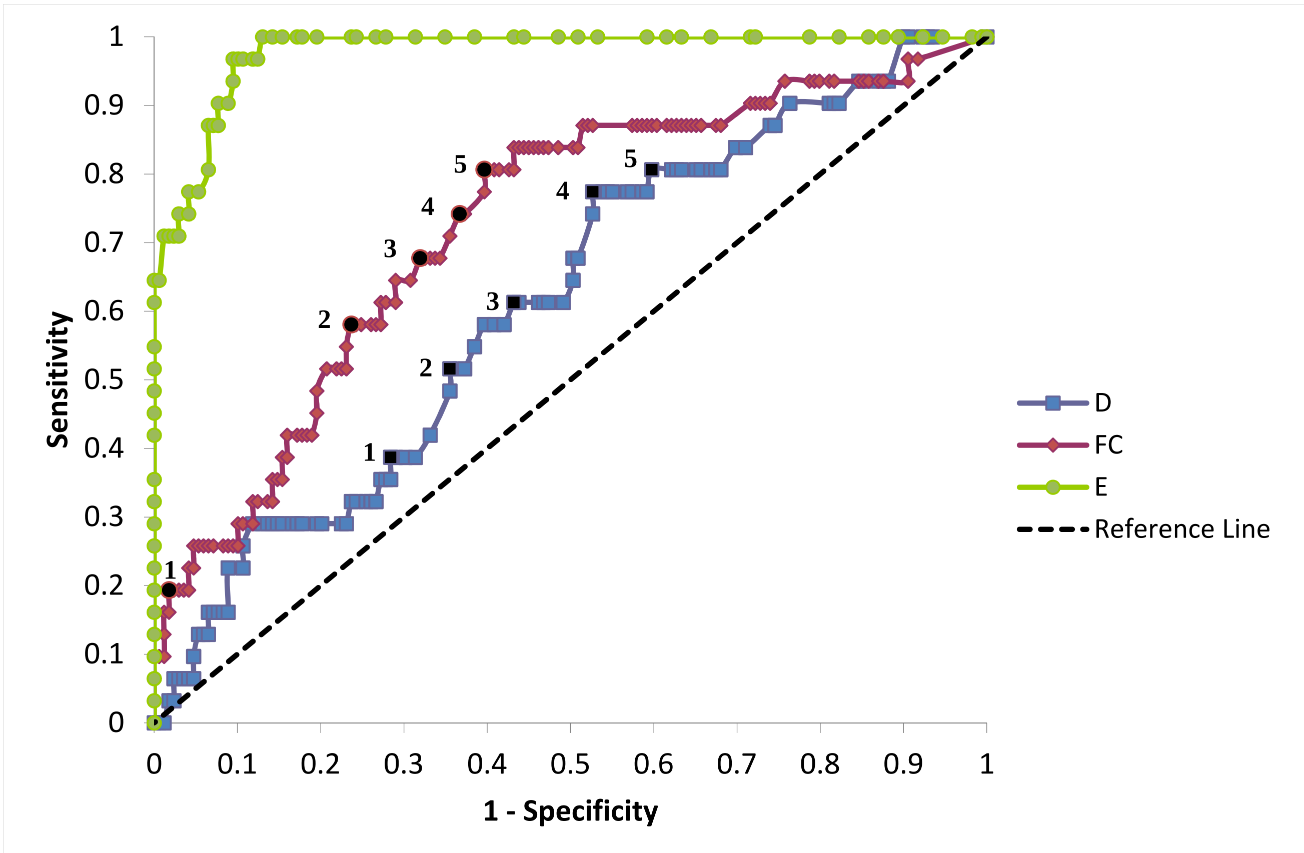


Figure 8. Receiver Operating Characteristic Curves for Three Rater Groups

Table 8 shows the results of using the classification algorithm (Figure 2) to classify the 200 drivers, based on the F/C ratings. The categorization used two threshold cutoffs, i.e., 55.69 and 72.53 to place drivers into three groups based on difficulty rating (basic, routine, accomplished) or into the group for drivers who data misfit the model (unable to categorize). Using the algorithm 23 of the 30 drivers who failed the on-test were identified as Group B (routine) or Group C (basic) and received a recommendation for at-risk drivers that emphasized the need follow-up with a physician and an occupational therapist for a CDE.

Table 8. Results of Using Algorithm for Categorizing Drivers

Group Name	*On-road test outcome*		*Critical Error*	*Subgroup 1*	*Subgroup 2*
Group A	**Fail**	**Pass**	**Critical Error**	**A1 Group**	**A2 Group**
N = 109	7	102	0	109	0
Group B	**Fail**	**Pass**	**Critical Error**	**B1 Group**	**B2 Group**
N = 73	17	56	6	67	6
Group C	**Fail**	**Pass**	**Critical Error**	**C1 Group**	**C2 Group**
N = 12	6	6	11	1	11
Group D	**Fail**	**Pass**	**Critical Error**	**C1 Group**	**C2 Group**
N = 6	1	5	2	4	2

Objective 3: Develop the instructional clinical outputs, or "keyforms."

Focus Group 1 (OTs)

Demographics. This focus group had twelve participants, ten women, and two men, five being occupational therapists and seven being OT/CDRSs. Job classifications were OT/CDRS in either community (n= 4) or academic setting (n=3), OT/Researcher (n=3), and OT/Administrative or Management (n=2).

Results. Data from the focus group questions were coded according to three themes: (1) Face Validity, (2) Appearance and Wording, and (3) Usability.

- Face Validity: Respondents said listing items hierarchically (easy to hard) by difficulty level improved face validity as did use of color coding to illustrate ratings and item hierarchy. They suggested we emphasize the area on the keyform where ratings change (e.g., transition zone where overall ratings shift from "a little difficulty" to "a moderate level of difficulty").
- Appearance and Wording: Formatting comments included that the keyform was too "busy" and "difficult to read." They suggested using a legend to clarify terms like "cautiously" or "dense traffic", using full items versus abbreviating, and increasing font size for "elder friendliness." Additionally, on results summary, show items without item number (in results they are listed by difficulty, not numerical order).
- Usability: Respondents said identifying the key domains where difficulty occurs (e.g., motor coordination) would help them match client deficits with occupational therapy interventions (e.g., vehicle modifications). The keyform may help identify driver limitations with potential to be addressed by the OT generalist before pursuing referral to a CDRS. The keyform could also help justify referral to and intervention by a CDRS. Suggestions for *revisions* were to allow space for the users to include comments, enable creation of reports comparing the different raters (e.g., driver vs. caregiver), and to incorporate training in use of the SDBM (e.g., case study), so users would better understand the results, driver profiles, and recommendations.

Focus Group 2 (Expert Panel)

Demographics. Five occupational therapists, all CDRSs, each with more than 10 years experience, participated. They were from four states representing the Southeast, Northeast, and Midwest. Four attended on-site and one via telephone conference.

Results. As illustrated in Table 1, the CDRSs perceived the SDBM as "a screening tool that can trigger conversations and broad decisions about driving", one that "measures behavior in such a way as to give caregivers a structured method of rating driving difficulty" and "allows information to be shared with the driver and professionals, such as a doctor or a CDRS." The keyform recommendations may enhance the clarity of communication about driving concerns, and increase the efficiency of an appointment scheduled with the doctor and/or CDRS to discuss driving-related issues. The CDRSs feedback on the 11 keyform questions (e.g., clarity, ease of use, readability, adequacy, understandability, and acceptability) are listed in Table 4-8 along with the mean VAS ratings of the expert panel members' responses. Using the VAS scale from "0" to "10", "0" was least acceptable while "10" was most acceptable. The overall average of the respondent's keyform ratings was 8.4, SD=0.8. Mean ratings ranged from 7.7-8.9, with the lowest rating given for Q10a – "How would you rate the acceptability of the keyform for drivers?" and the highest rating for Q5 – "Does the keyform adequately illustrate the transition zone, i.e., where the ratings shift from "No Difficulty" to "A Little Difficulty"?"

Table 9. Focus Group 2 (Expert Panel): Keyform Questions and Select Respondent Data

Questions	*Mean rating +SD*	*Respondent comments*
Q1. From the case studies – does the keyform adequately demonstrate the differences in drivers' abilities?	8.1 +1.8	- caregiver report remarkably in line with the therapist's measure of abilities (R1) - easy to compare good/marginal/bad (R2) - yes, very clear, colors help (R4)
Q2. How would you rate the ease of use of the keyform?	8.3 +1.5	- impressed with ease of getting a visible snapshot of the abilities (R1) - shows great promise in ease of use and understandability (R1) - user might not understand how overall score derived from ratings (R2)
Q3. How would you rate the clarity of the item hierarchy?	8.2 +1.0	- hierarchy helps client / family understand that despite many abilities intact, impaired critical elements lead to results/recommendations (R1)
Q4. Does the keyform adequately illustrate the driver's areas of difficulty?	7.9 +1.7	- caregiver self-report was impressively consistent to therapist's rating (R1) - yes, the colors are great! (R2) - colored zones are great (R5)
Q5. Does the keyform adequately illustrate the transition zone, i.e., where the ratings shift from "No Difficulty" to "A Little Difficulty"?	9.4+0.7	- Yes, very understandably (R1) - Excellent! (R2) - Colors very helpful (R5)
Q6. How would you rate the readability (font, spacing, orientation) of the keyform?	8.8 +0.9	- once oriented, I found it clear (R2) - positive value that Web-based version will offer further description (R1) - excellent! (R2)
Q7. How would you rate the understandability of the language used to describe the items?	7.9 +1.7	- clearly seems on the right track (R1) - great (R2) - some items need clarification or specific examples (R4)
Q8. How would you rate the acceptability of the keyform layout?	8.9 +0.9	- once oriented I found it easier (R1) - great (R2) - the hierarchy is ideal and enables someone to educate on driving environment and driving situations (R3) - excellent (R5)
Q9. How would you rate the acceptability of the keyform for occupational therapists?	8.8 +1.2	- once understood by OTs would be very eagerly accepted (R1) - great visual when talking to patients/family (R3) - provide instructions prior to using (R4) - useful (R5)
Q10a. How would you rate the acceptability of the keyform for drivers?	7.7 +1.5	- builds self-awareness of deficits (R1) - might only relate to colors and average score especially if they have rated themselves 5s (not difficult) (R2) - provide instructions (R4) - explain the layout / meaning (R4)
Q10b. How would you rate the acceptability of the keyform for caregivers?	8.2 +1.2	- could strongly enhance the therapeutic discussion(R1) - provides rationale for restriction or cessation (R1) - should definitely trigger conversation (R2) - provide instructions (R4) - explain the layout / meaning (R4)
Overall mean and SD	8.4 +0.8	_

Legend: Q= Question; R= Respondent; Not all raters provided written responses for feedback, SD= Standard deviation

Note: * Numerical data derived from the Visual Analogue Scale are used as continuous data.

The expert panel helped us operationalize three driver types or profiles ("pass, borderline, fail"), and for each profile identify specific safety needs for continued driving (or driving cessation) and the logical next steps for F/C. Panelists discussed the clinical, ethical, and legal implications of making recommendations, and sought the "just right fit recommendation" for each driver profile ("pass, borderline, fail").

For the most impaired driver groups ("borderline" or "fail"), they were concerned that an overly severe rating may lead to caregiver-driver conflict, that the caregiver would unnecessarily "take the driver off the road", or "reject the screening results". On the other hand, they felt lenient recommendations may prevent F/C of at-risk drivers from taking appropriate steps to improve safety. Respondents suggested recommendation language that would facilitate action while minimizing negative impact of words pertaining to "threat" "risk" and "concern", which resulted in much debate. One respondent suggested easing the negative impact of a recommendation by "starting with the good", and highlighting items (driving behaviors) that the driver was able to perform, and not just pointing out areas of difficulty.

For the best driver in the group ("passed") the expert panel members developed the following description, refined by the research team, to be shared with the F/C:
Category: Accomplished Driver- Driving is overall good, but difficulty is experienced with some challenging driving situation, e.g., (examples are selected from the driver's profile). Recommendation: It may be helpful to avoid or limit the challenging driving situations (described in the example). Based on your ratings, we do not think that a comprehensive driving evaluation is critical at this time; but we recommend completing this screening at least annually or if there are any changes in the driver's status.

The expert panel members also suggested specific recommendations for the "borderline" or "fail" driver profiles including recommendation to have a comprehensive driving evaluation by an OT with specialty certification. They also suggested general recommendations for all groups such as: "as suggested by the American Geriatrics Society seek a physical and eye exam annually or earlier" or "take a mature drivers class offered by AAA or AARP".

Focus Group 3 (F/C)
Demographics. Seven respondents included five spouses (71.4%), one adult child (14.3%), and one friend (14.3%). Age range was 46-77 years (median age= 65); most were females (57.1%); 42.9% Caucasian (n=3), 28.6% African-American (n=2), and 28.6% Asian (n=2); all had at least high school graduation, with most having a Bachelor's or higher degree (57.1%).

Results. Changes were recommended for both the Web-based SDBM and the keyform. Changes included: to rename "caregiver" as "proxy" which indicated (more accurately) a family member, friend or caregiver with sufficient knowledge to rate the driver's ability; revise instructional scripts for the Web-based SDBM; and incorporate "drop down boxes" to document numerical values e.g., birth year. They suggested clarify the race question (SDBM Section A-demographics); create a proxy version of the driving history (SDBM Section B); and consider use of "not applicable" versus forced response for the driving behavior questions (68 items of SDBM Section C). Respondents also requested a customer satisfaction survey be included with the Web-based SDBM and keyform. Table 4-9 presents the F/C VAS ratings on the six questions regarding purpose, clarity, understandability and meaningfulness of the Web-based keyform. The mean VAS score across raters was

9.01/10 and the SD=1.02.

Table 10. Focus Group 3: F/C Visual Analogue Scale Ratings*

	Rater A	*Rater B*	*Rater C*	*Rater D*	*Rater E*	*Rater F*	*Rater G*	*Mean of Sum*	*SD of Sum*
Q1a. How well did we explain the purpose of the questionnaire?	8.4	8.7	8.1	9.8	9.9	9.9	10	9.26	0.82
Q1b. How clear were the instructions of the questionnaire?	6.8	8.4	8.1	9.7	6.3	9.8	7.7	8.11	1.33
Q2a. How well did we explain the purpose of the keyform?	7.6	8.4	9.1	10	9.5	9.8	9.9	9.19	0.89
Q2b. Is the keyform useful, e.g., does it illustrate your areas of concern ?	8.8	8.3	9.4	10	9.7	9.8	9.9	9.41	0.64
Q2c. Is the keyform understandable, e.g., does it reflect the driver's difficulties?	8.3	8.1	7.5	9.9	7.6	9.7	10	8.73	1.10
Q2d. Is the keyform meaningful, e.g., does it provide helpful recommendations regarding follow-up?	7.5	9.1	9.4	9.9	9.7	9.9	10	9.36	0.88
Mean of Sum	7.90	8.50	8.60	9.88	8.78	9.82	9.58	9.01	--
SD of Sum	0.73	0.35	0.80	0.12	1.48	0.08	0.92	--	1.02

Note: * Data derived from the Visual Analogue Scale are used as continuous data.

Objective 4: Develop a data collection system

The team was successful in establishing a Web-based SDBM with a data collection system utilizing the secure computer network in the University of Florida – College of Public Health and Health Professions. The Web-based measure was reviewed and piloted by team members and peers. Comments captured from peer and team reviews were utilized to inform revisions to the Web-based measure. The UF database allows us to capture and store SDBM ratings and to track users while protecting rater confidentiality. The Web-based measure, now named the Fitness-to-Drive Screening Measure (FTDS), has an on-line user manual, and functions for end users to retrieve keyforms and recommendations following completion of the FTDS.

CHAPTER 5 - DISCUSSION

Based on our findings from objectives 1 through 4, we were able to establish psychometric properties in support of the SDBM. Furthermore, refinements based on those findings supported the creation of a Web-based measure with useable outputs for older drivers, F/C and occupational therapists.

Task 1(a): Confirm the factor structure of the SDBM.
Our findings demonstrated appropriate data fit to the one factor model following revision to the 54 item SDBM. These revisions were supported by knowledge of the dimensionality and factor loading of the items to two factors (pre-driving and driving skills) based on initial testing. Driving items were retained in the Section C and rated for difficulty while other items were moved to the Section B driving history, improving the measurement precision of the SDBM.

Task 1(b): Describe the item and person level properties of the SDBM.
We investigated the psychometric properties of the 68-item SDBM by unidimensionality and local independency, rating scale, item/person-level psychometrics, and item hierarchy across three groups (older drivers, F/C and driving evaluators).

We tested 200 older drivers who were mainly white with the majority having college degrees, relatively healthy and cognitively intact. Likewise, the 200 F/C were mainly white, but just less than half of the group had college degrees and the majority was female. Interestingly, although all of the F/C were licensed drivers, a third stated that their independence would be impacted if their spouse/partner, the older driver, stops driving. Bias may play a role in the caregivers' reports in two ways: first, by showing concern for their loved one's safety, and second, being more concerned with maintaining their own means of transportation, which was likely to occur in the group that stated that they will be impacted when their spouse/partner stops driving.

The rating scale structure suggested that the "Cannot Do" category was under used across three rater groups, therefore we collapsed the "Cannot Do" and "Very Difficult" categories.

Item/ person-level psychometrics of the SDBM for each of the three groups revealed incongruence pertaining to (mis)fit. This overlapping misfitting item 38 "Use a map while driving" in the driver and caregiver groups may need clarification as group members were not specifically instructed on the *type* of map (Google map, or a geographic positioning system map), which could lead to greater variability in their response choices. Misfitting items (18.52% using criterion A, 9.26% using criterion B） in the evaluator group were problematic.

Across the three rater groups, these data displayed good *person separation* (>3.07) and *item separation* (>5.43), *good item reliability* (>0.97) and *person reliability* (>0.90) and *Cronbach's alpha* >96%. However, some of the items did not follow the hypothesized order of item difficulty. The evaluator group's ratings showed a different item hierarchy compared to the other two rater groups, potentially due to the evaluators wanting to minimize traffic risk and maximize participant safety.

Even though mild ceiling effects existed for the F/C (14%) and driver (6.5%) groups, and the person mean across the three rater groups was about 2 SD higher than the item mean, given that this sample was high functioning, the SDBM may have a sufficient level of challenging items to measure other older adult groups.

Limitations: Caution needs to be exercised when interpreting the data as we can only generalize results to the sample under this study. Several pairs of easy items showed local dependency and some of them were misfitting as well. This issues were partially addressed by revisions to the sections of the SDBM (e.g., moving some items to the driving history section where difficulty ratings is not scored).

These findings indicated significant differences between the ratings of the evaluator and the caregiver on 19 items, where the evaluator rated 9 items more severely compared to the caregiver; and the caregiver rated 10 items more severely compared to the evaluator. There were no significant differences between the ratings of the evaluator and the driver; or the caregiver and the driver. A user manual provided with the Web-based SDBM provides a training video for F/C that will be used to instruct them on rating guidelines and potentially improve rating accuracy.

Our data reflect the SDBM is efficient and can accurately classify a population of older drivers with varying ability levels into distinct groups with more or less of safe driving behaviors; as such, the SDBM, may provide the first step to identify unsafe driving behaviors and provide occupational therapists with an entry point for delivering preventative services.

Task 1(c): Determine the rater severity of the three rater groups

Inter-rater reliability

We found statistically significant correlation among the ratings of the driver, evaluator, and the F/C groups. Although the correlation was low between driver and evaluator group, a strong correlation existed between F/C and evaluator group.

Rater Effects

Facet ruler of the SDBM. The distribution of the drivers' ability relative to the distribution of the items' difficulty indicates that the participants in this study performed well on the instrument. As can be seen from Figure 4-4 many of our items are on the same logit level. Taking into account that only the *means* of the items are represented, we have more overlapping among the items because each item consists of four difficulty levels corresponding to a 4-point adjectival scale. Having different items at the same difficulty level in the item pool may be redundant for paper and pencil tests; however, that will increase the item pool, which will in turn provide more choices for future applications, such as using computer adaptive testing (the next step in the development of our instrument).

Fit statistics of the three rater groups. The fit statistics across the rater groups (evaluators and F/C) showed that there were no erratic rater groups and that the evaluator group had overall more *severe* raters (Facets) when compared to driver and F/C group.

Fixed chi-square and paired comparisons. While the evaluator group was a more severe rater group than the

F/C group on the overall scale, the F/C group rated ten items more severely than the evaluator group. On the other hand the evaluator group rated nine items more severely than the F/C group. Evaluators have formal training to rate driving behaviors according to the standards of regulatory bodies such as the Department of Motor Vehicle and Highway Safety licensure guidelines; and we can therefore expect that they will be more technical and more stringent in their ratings. The F/C group does not have such formal training and are rating the drivers on their perceptions of how they are experiencing the driving safety of their loved ones. The tendency for evaluators to rate more severely (than the F/C) may be influenced by their training to focus on identifying deficits. The F/Cs, on the other hand, may be influenced by showing concern for their loved one's safety, thus rating more stringently, or being concerned with maintaining their own independence in transportation and rating leniently.

The generalizability of our findings are limited due to using only two evaluators and the sample characteristics: white (88.5%) and educated drivers (57.0% had some college education or university degree), and the F/C being mainly female (72.5%), white (90.0%), and having completed college or university degrees (46.5%).

Due to the significant relationship between F/C and evaluator findings, we will produce short educational videos to train the caregivers to recognize older adults' unsafe driving behaviors more precisely.

Objective 2. We validated the SDBM to results from on-road driving evaluation.

The area under the curve (AUC) of 0.513 based on Drivers' SDBM rating, although statistically significant, yielded low accuracy in predicting the on-road driving test results. This poor predictive ability is also evident when examining the SDBM sensitivity, specificity, PPV, and NPV when based on the drivers' ratings (see Table 6). As such, we conclude that the SDBM, when used by drivers, is not an accurate self-report screening tool to make determinations regarding on-road outcomes. That being said, the driver's ratings may still be used by occupational therapists in discussing differences between drivers' self-ratings and those of the F/C to increase self-awareness of driving behaviors. Likewise, the driver report may also be used, in combination with the caregiver's report to "start" the conversation about future driving interventions, driving alternatives, or driving cessation.

The F/Cs' AUC of 0.726 yielded acceptable (greater than 0.7) accuracy for using the SDBM measure to predict outcomes of the on-road driving test. Based on the F/C ROC and Table 7 data of sensitivity, specificity, PPV and NPV the optimal Logit score for classifying drivers based on F/C rating was between cut-point 4 (71.92) and cut-point 5 (73.47), this knowledge helped us set our upper threshold for driver categorization at 72.53. Drivers who fell below this threshold (the basic and routine drivers) received a recommendation which included seeking a doctor's appointment and pursuing a CDE with an occupational therapist.

To our knowledge, this is the first screening tool showing concurrent criterion validity based on F/C ratings classifying older drivers who failed an on-road test. As such, occupational therapists may use this screening tool (completed by F/C) to form a picture of the driver's driving behaviors. This screening tool may also be used to facilitate a conversation about difficulty with driving (from the caregiver and/or client perspective), and help to identify driving problems, which may in turn lay the foundation for intervention planning by a certified driver rehabilitation specialist or evaluator. Moreover, the SDBM operationalizes driving by means of 54 behavioral items. Thus it gives the practitioner, perhaps a generalist who is not extensively familiar with all the underlying driving-related issues, a concrete description of driving abilities that can be viewed as “difficult” to perform, and provide an entry point for clinical decision-making, intervention, adaptation (e.g., suggesting safer strategies, such as not driving on the interstate) or referral to a driving rehabilitation specialist.

Limitations beyond those already mentioned (e.g., race) pertain to the error associated with the family members/caregivers SDBM ratings, the less than desirable specificity, low PPV, and only including two sites in the testing of participants.

Objective 3: Develop the instructional clinical outputs, or “keyforms”
The OTs’ results supported the Web-based SDBM and keyform as a potentially useful tool to provide a profile of the driver for further decision-making. It may facilitate communication about driving difficulties among the stakeholders and the drivers. Velozo & Woodbury (2011) suggested that a major benefit of the keyform is that it can be used as the basis for interventions. In our focus group, the OTs have verified the usefulness of the keyform to “provide a visible snapshot of abilities” from which further interventions could be planned.

Based on the expert panel of CDRSs’ specialized knowledge, in-depth understanding and clinical reasoning (American Occupational Therapy Association, 2010), we refined the classifications of drivers. This led to formulating the “just right fit” recommendations for three driver profiles, with wording and action steps to guide F/C, and potentially promote safety among drivers.

From the F/C feedback we learned the Web-based SDBM and keyform were useful to rate and share a driver’s ability level amongst the driver, F/C, or health care providers. We implemented their suggestions to enhance the functionality, user-friendliness, understandability, and acceptability of the Web-based version of these tools.

Limitations of this study pertain to generalizability of the results, which can only be extrapolated to participants fitting the profile of our respondents. However, we used purposive sampling for this study which yielded reasonable representation of participants i.e., occupational therapists from clinical and academic settings; experts from three U.S. states and different practice settings, and F/C from different age, gender, and racial groups.

A key strength was our inclusion of three stakeholder groups who shared their specific needs, perspectives and suggestions and enhanced the Web-based SDBM and keyforms. The respondents’ descriptions and input reflected the many aspects of instrument development essential to satisfy the needs of different user groups (Thurstone, 1925). To our knowledge, this is the first study to include OTs, expert CDRSs, and F/C, in refining a driving outcome measure.

Objective 4: Develop a data collection system
The measure was renamed the Fitness-to-Drive Screening measure (FTDS). The FTDS utilizes a four-point rating scale and the number of items for which difficulty is scored (Section C) was reduced to 54 items. Based on positive team and peer review and piloting we are initiating dissemination of the FTDS at state and national

levels through collaborating organizations such as the Safe Mobility for Life coalition. The Web-based FTDS database will provide data on user ratings, demographics, and usage patterns that will facilitate future research. We will continue to obtain feedback via a Web-based user survey on the FTDS features (rating process, on-line user manual, keyforms, and recommendations) to inform future revisions.

CHAPTER 6 - CONCLUSION

Based on our findings, we assert the SDBM may be useful for: (1) family members and caregivers to identify at-risk older drivers and to follow logical next steps based on keyform recommendations; (2) occupational therapy practitioners to identify an entry point for further interventions or referrals; and (3) Certified Driver Rehabilitation Specialists (CDRS) to develop realistic and targeted intervention goals to promote driving fitness. Next steps are to test the Web-based version of the Fitness-to-Drive Screening Measure via additional web-sites. This will provide further data on the adequacy, readability, acceptability, and usability of Web-based version. Further testing will also support our goals to help family members and caregivers identify at-risk older drivers while supporting the roles of occupational therapy generalists and CDRSs as clinicians and advocates for increased driving safety of older adults. Future implications for this work, from the perspective of our FDOT partners, are numerous. Even though not empirically tested, we believe that this tool may have widespread use in the State of Florida, and beyond, as expressed by the following:

- We believe that this tool will have applicability to be used among other health care practitioners. For example, we envision that primary care physicians, who must make determination on fitness-to-drive, may use this on-line tool in their waiting rooms. As such, this standardized tool may be completed by a loved one, family member or caregiver who are accompanying the “driver” and who have observed this person’s driving behavior in the last three months. The web-based FTDS provides a driver profile, driver classification and recommendations which the physician may print. This information may enhance readiness to make determinations on fitness-to-drive, rather than just relying on clinical reasoning alone.
- Although not empirically tested, we believe that this tool may be useful for agents of the aging network (e.g., AARP mature driver operators, AAoA affiliates, or AAA clubs). Such delegates, may use this tool as a starting point for conversations pertaining to fitness-to-drive; and they may use the profiles to underscore recommendations for follow-up assessments.
- The stakeholders of the Florida Safe Mobility for Life Coalition must be informed on the psychometrics, usability and utility of this tool to enhance the action plans that were set and identified by a number of the emphasis areas, e.g., the Assessment, Remediation, and Rehabilitation, and/or Prevention and Early Recognition emphasis areas.
- The FDOT State Safety Office is the lead in the current revision to the Strategic Highway Safety Plan (SHSP). In this revision, the Safe Mobility for Life Coalition’s Aging Road User Strategic Safety Plan will now be incorporated into the new At-Risk Drivers Emphasis Area. This change has the potential to provide further outreach and education of the FTDS to additional SHSP stakeholders.

The University of Florida (UF) will continue to seek support to “test” the usability of this tool among various stakeholders including, but not limited to:

- Medical community (e.g., primary care physicians)
- National, state, and local organizations (e.g., American Occupational Therapy)
- Diverse end-users (e.g. social workers)

With these recommendations, UF researchers and our FDOT partners at the State Traffic Engineering and Operations Office will be able to confidently disseminate this tool as one that is valid and reliable among all identified stakeholders, organizations, and end-users.

REFERENCES

American Occupational Therapy Association (2010). *Statements: Driving and community mobility*. Bethesda, MD: American Occupational Therapy Association.

Ball, K. K., & Owsley, C. (1993). The useful field of view test: a new technique for evaluating age- related declines in visual function. *Journal of the American Optometrics Association, 64*(1), 71-79.

Bédard, M., Porter, M. M., Marshall, S., Isherwood, I., Riendeau, J., Weaver, B., Tuokko, H., Molnar, F., & Miller-Polgar, J. (2008). The combination of two training approaches to improve older adults' driving safety. *Traffic Injury Prevention, 9*(1), 70-76.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: fundamental measurement in the human sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. NY: Guilford Press.

Classen, S., Horgas, A., Awadzi, K., Messinger-Rapport, B., Shechtman, O., & Joo, Y. (2008). Clinical predictors of older driver performance on a standardized road test. *Traffic Injury Prevention, 9*, 456-462.

Classen, S., McCarthy, D. P., Shechtman, O., Awadzi, K. D., Lanford, D. N., Okun, M. S., Rodriguez, R. L., Romrell, J., Bridges, S., Kluger, B., & Fernandez, H. H. (2009). Useful Field of View as a reliable screening measure of driving performance in people with Parkinson's disease: Results of a pilot study. *Traffic Injury Prevention, 10*(6), 593 – 598.

Classen, S., Shechtman, O., Awadzi, K. D., Joo, Y., & Lanford, D. N. (2010). Traffic violations versus driving errors of older adults: Informing clinical practice. *American Journal of Occupational Therapy, 64*(2), 233-241.

Classen, S., Wang, Y., Velozo, C. A., Bédard, M., Winter, S. M., & Lanford, D. N. (in press). Concurrent criterion validity of the Safe Driving Behavior Measure: A predictor of on-road driving outcomes. *American Journal of Occupational Therapy*.

Classen, S., Wen, P. -S., Velozo, C. A., Bédard, M., Winter, S. M., Brumback, B., & Lanford, D. N. (2012a). Psychometrics of the self-report Safe Driving Behavior Measure for older adults. *American Journal of Occupational Therapy*.

Classen, S., Wen, P. -S., Velozo, C. A., Bédard, M., Winter, S. M., Brumback, B. A., & Lanford, D. N. (2012b). Rater reliability and rater effects of the Safe Driving Behavior Measure. *American Journal of Occupational Therapy, 66*(1), 69-77.

Classen, S., Winter, S. M., Velozo, C. A., Bédard, M., Lanford, D., Brumback, B., & Lutz, B. J. (2010). Item development and validity testing for a Safe Driving Behavior Measure. *American Journal of Occupational Therapy, 64(2)*, 296-305.

Croston, J., Meuser, T. M., Berg-Weger, M., Grant, E. A., & Carr, D. B. (2009). Driving Retirement in Older Adults with Dementia. *Topics in Geriatric Rehabilitation, 25*(2), 154-162.

Edwards, J. D., Ross, L. A., Wadley, V. G., Clay, O. J., Crowe, M., Roenker, D. L., & Ball, K. K. (2006). The useful field of view test: normative data for older adults. *Archives of Clinical Neuropsychology, 21*(4), 275-286.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288.

Jette, A. M., & Haley, S. M. (2005). Contemporary measurement techniques for rehabilitation outcomes assessment. *J Rehabil Med, 37*(6), 339-345.

Jette, A. M., Tao, W., Norweg, A., & Haley, S. (2007). Interpreting rehabilitation outcome measurements. Journal of Rehabilitation Medicine, 39(8), 6.

Justiss, M. D., Mann, W. C., Stav, W., & Velozo, C. (2006). Development of a Behind-the-Wheel Driving Performance Assessment for Older Adults. *Topics in Geriatric Rehabilitation, 22*(2), 121-128.

Kay, L. G., Bundy, A. C., & Clemson, L. M. (2008). Predicting fitness to drive using the visual recognition slide test (USyd). *American Journal of Occupational Therapy, 62*(2), 187-197.

Kay, L. G., Bundy, A. C., & Clemson, L. M. (2009). Predicting fitness to drive in people with cognitive impairments by using DriveSafe and DriveAware. *Archives of Physical Medicine and Rehabilitation, 90*(9), 1514-1522.

Kay, L. G., Bundy, A. C., Clemson, L. M., & Jolly, N. (2008). Validity and reliability of the on-road driving assessment with senior drivers. *Accident Analysis & Prevention*, *40*(2), 751-759.

Kielhofner, G., Dobria, L., Forsyth, K., & Basu, S. (2005). The construction of keyforms for obtaining instantaneous measures from the Occupational Performance History Interview rating scales: Empirical quantitative study *OTJR: Occupation, Participation and Health, 25*(1), 23-32.

Krueger, R. A. (2009). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: SAGE Publications.

Linacre, J. M. (2000). Comparing "Partical Credit" and "Rating Scale" models. *Transactions, 14*(3), 768.

Linacre, J. M. (2002a). Understanding Rasch Measurement: Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Linacre, J. M. (2002b). What do Infit and Outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Linacre, J. M. (2004). Facets Rasch measurement computer program. Chicago: Winsteps.com.

Linacre, J. M. (2012a). Winsteps® (Version 3.74.0). Beaverton, OR: Winsteps.com.

Linacre, J. M. (2012b). *Winsteps® Rasch measurement computer program User's Guide*. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2012) Facets computer program for many-facet Rasch measurement, version 3.70.0. Beaverton, Oregon: Winsteps.com

Magasi, S., & Heinemann, A. W. (2009). Integrating stakeholder perspectives in outcome measurement. *Neuropsychological Rehabilitation, 19*(6), 928-940.

Marottoli, R. A., Cooney, L. M., Jr., Wagner, R., Doucette, J., & Tinetti, M. E. (1994). Predictors of automobile crashes and moving violations among elderly drivers. *Annals of Internal Medicine, 121*(11), 842-846.

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

Morse, J. M. (1994). Designing Funded Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 220-235). Thousand Oaks, CA: SAGE publications.

Muthen, L. K., & Muthen, B. O. (2006). *Mplus User's Guide* Los Angeles, CA: Muthen & Muthen.

Myers, A. M., Paradis, J. A., & Blanchard, R. A. (2008). Conceptualizing and measuring confidence in older drivers: Development of the day and night driving comfort scales. *Archives of Physical Medicine & Rehabilitation, 89*(4), 630-640.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement, 4*(4), 386-422.

Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement, 5*(2), 189-227.

National Alliance for Caregiving, & AARP (2009). *Caregiving in the U.S.: A focused look at those caring for the 50+*. Bethesda, MD: National Alliance for Caregiving and MetLife. Retrieved Oct 1,2011, from http://assets.aarp.org/rgcenter/il/caregiving_09.pdf

NVivo qualitative data analysis software (2008) (Version 8). Doncaster, Victoria, Australia: QSR International Pty Ltd.

Patomella, A. H., Kottorp, A., & Tham, K. (2008). Awareness of driving disability in people with stroke tested in a simulator. *Scandinavian Journal of Occupational Therapy, 15*(3), 184-192.

Patomella, A. H., Tham, K., & Kottorp, A. (2006). P-drive: assessment of driving performance after stroke. *Journal of Rehabilitation Medicine, 38*(5), 273-279.

Portney, L., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall Health.

Posse, C., McCarthy, D. P., & Mann, W. C. (2006). A pilot study of interrater reliability of the Assessment of Driving-related Skills: Older driver screening tool. *Topics in Geriatric Rehabilitation, 22*(2), 113-120.

Shechtman, O., Classen, S., Awadzi, K. D., & Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: A validation study. *Traffic Injury Prevention, 10*(4), 379-385.

Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal of Quality Health Care, 18*(5), 377-382.

SPSS Inc. (2012). PASW Statistics (Version 20). Chicago, IL: SPSS Inc.

Stav, W. B., Justiss, M. D., McCarthy, D. P., Mann, W. C., & Lanford, D. N. (2008). Predictability of clinical assessments for driving performance. *Journal of Safety Research, 39*(1), 1-7.

Stereo Optical Company Inc. (2007). *Reference and instruction manual: Optec 2500 Vision Tester*. Chicago, Illinois.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to Receiver Operating Characteristics curves. *Canadian Journal of Psychiatry, 52*(2), 121-128.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th ed.). Oxford: Oxford University Press

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology, 16*(7), 433-451.

Velozo, C. A., & Woodbury, M. L. (2011). Translating measurement findings into rehabilitation practice: An example using Fugl-Meyer Assessment-Upper Extremity with patients following stroke. *Journal of Rehabilitation Research and Development, 48*(10), 1-11.

Wang, W.-C., & Chen, C.-T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. *Educational and Psychological Measurement, 65*, 376-404

Wild, K., & Cotrell, V. (2003). Identifying driving impairment in Alzheimer disease: a comparison of self and observer reports versus driving evaluation. *Alzheimer Disease and Associated Disorders, 17*(1), 27-34.

Winter, S. M., Classen, S., Bédard, M., Lutz, B., Velozo, C. A., Lanford, D. N., & Brumback, B. A. (2011). Focus group findings for the self-report Safe Driving Behavior Measure. *Canadian Journal of Occupational Therapy*, *78*(2), 72-79.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: Mesa Press.

FTDS Version II
6/13/2012

FITNESS-TO-DRIVE SCREENING MEASURE

A.1 Demographics of Rater

Instructions:

1. **Please answer all questions to the best of your ability.**
2. **In this section we are asking questions about you, the family member, friend or caregiver who is rating the driver.**
3. **Answer by checking the box or filling in the blank.**

1. What is your birth year? _______

2. What is your gender?

- □ Male
- □ Female

3. What is your ethnicity? Do you consider yourself to be:

- □ **Hispanic or Latino** (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race)
- □ **Not Hispanic or Latino**

4. What is your race? Would you say you are: (Choose one)

- □ **American Indian / Alaska Native / First Nations / Aboriginal or Inuit**: having origins in any of the original peoples of North, Central, or South America, and who maintains tribal affiliation or community attachment.
- □ **Asian**: having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- □ **Black or African American**: having origins in any of the black racial groups of Africa.
- □ **Native Hawaiian or Other Pacific Islander**: having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
- □ **White**: having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- □ Other: specify ___________

5. What is your highest level of education?

- □ Did not go to school
- □ Completed Grade school (5th grade)
- □ Completed Middle school (8th grade)
- □ Completed High School/G.E.D. (12th grade)
- □ Completed Vocational Training
- □ Some College after High School Graduation
- □ Associate Degree
- □ Bachelor’s Degree
- □ Some Professional School after College Graduation
- □ Master’s Degree
- □ Doctoral Degree

6. Do you have a driver’s license?

- □ No
- □ Yes

7. How many days a week do you typically drive?

- □ Less than 1
- □ 1
- □ 2
- □ 3
- □ 4
- □ 5
- □ 6
- □ 7

8. Do you live alone? (If “Yes” – Go to question # 10)

- □ No
- □ Yes
- □ Mostly (for part of the year)

9. Who lives with you?

- □ Spouse or partner
- □ Child
- □ Family/Other relative: specify: _________
- □ Friend(s)
- □ Paid caregiver
- □ Other: specify ______________________

10. What is your relationship with the driver you are rating?
    - ◻ Spouse or partner
    - ◻ Child
    - ◻ Family/Other relative: specify: ________
    - ◻ Friend(s)
    - ◻ Paid caregiver
    - ◻ Other: specify __________________

11. Besides you, how many other licensed drivers are in your household? _______

12. Do you rely on the driver for any of the following trips or activities? (Choose all that apply)
    - ◻ Shopping
    - ◻ Grocery store
    - ◻ Social activities
    - ◻ See friends or family
    - ◻ Church
    - ◻ See doctor or get medical care
    - ◻ Work related activities
    - ◻ Other (please list)______________________________
    - ◻ Do not rely

13. How many days a week do you ride with the driver you are rating?
    - ◻ Less than 1
    - ◻ 1
    - ◻ 2
    - ◻ 3
    - ◻ 4
    - ◻ 5
    - ◻ 6
    - ◻ 7

14. If the driver you are rating reduced or stopped driving, would it significantly impact your current lifestyle?
    - ◻ No
    - ◻ Yes

15. If “Yes” to question 14, please explain___________________________

FTDS Version II
6/13/2012

Fitness-to-Drive Screening Measure

A.2 Demographics of Driver

Instructions:

4. **Please answer all questions to the best of your ability.**
5. **In this section, we are asking questions about the driver you are rating.**
6. **Answer by checking the box or filling in the blank.**

1. What is the driver's birth year? _______

2. What is the driver's gender?
   - □ Male
   - □ Female

3. What is the driver's ethnicity? Do you consider the driver to be:
   - □ **Hispanic or Latino** (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race)
   - □ **Not Hispanic or Latino**

4. What is the driver's race?
   - □ **American Indian / Alaska Native / First Nations / Aboriginal or Inuit**: having origins in any of the original peoples of North, Central, or South America, and who maintains tribal affiliation or community attachment.
   - □ **Asian**: having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
   - □ **Black or African American**: having origins in any of the black racial groups of Africa.
   - □ **Native Hawaiian or Other Pacific Islander**: having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
   - □ **White**: having origins in any of the original peoples of Europe, the Middle East, or North Africa.
   - □ Other: please specify ___________

5. Does the driver live alone? (If “Yes” – Go to question #8)

- □ No
- □ Yes
- □ Mostly (for part of the year)

6. Who lives with the driver?

- □ Spouse or partner
- □ Child
- □ Family/Other relative
- □ Friend(s)
- □ Paid caregiver
- □ Other: please specify __________________

7. Besides the driver you are rating, how many other licensed drivers are in his/her household? _______

8. What is the driver’s highest level of education?

- □ Did not go to school
- □ Completed Grade school (5th grade)
- □ Completed Middle school (8th grade)
- □ Completed High School/G.E.D. (12th grade)
- □ Completed Vocational Training
- □ Some College after High School Graduation
- □ Associate Degree
- □ Bachelor’s Degree
- □ Some Professional School after College Graduation
- □ Master’s Degree
- □ Doctoral Degree

9. Does the driver use any of the following assistive devices? (Choose all that apply)

- □ Corrective lenses (such as eyeglasses or contacts)
- □ Hearing device \ hearing aid
- □ Mobility device (such as cane, walker, wheelchair)
- □ Car devices (such as seat pad, pedal assist, spinner knob)
- □ Other: please specify_____________________________

Fitness-to-Drive Screening Measure

FTDS Version II
6/13/2012

B. Driving History Profile

Instructions:

7. **Please answer all 31 questions to the best of your ability.**
8. **Answer by checking the box or filling in the blank.**

DRIVING HISTORY

1. How many days a week does the driver typically drive?
   - □ Less than 1
   - □ 1
   - □ 2
   - □ 3
   - □ 4
   - □ 5
   - □ 6
   - □ 7

2. Who usually rides with the driver? (Check all that apply)
   - □ Spouse / Partner
   - □ Family member/ Other relative
   - □ Friend
   - □ Paid caregiver
   - □ No one
   - □ Other (list)______________________

3. Has a health condition limited the driver's ability to drive?
   - □ No
   - □ Yes

4. Has taking medications limited the driver's ability to drive (over the counter or prescribed)?
   - □ No
   - □ Yes

5. Did the driver get any of the following tested in the last year? (Check all that apply)

- ◻ Vision
- ◻ Hearing
- ◻ Physical exam / checkup
- ◻ Other tests (list)_______________________________________

6. In the past year, did the driver complete/ or have done, any of the following car maintenance? (Check all that apply)

- ◻ Oil change
- ◻ Checking tires
- ◻ Checking fluid levels
- ◻ Checking headlights, brake lights and parking lights

7. Does the driver avoid (when possible) any of these driving situations? (Check all that apply)

- ◻ Rush hour/heavy traffic
- ◻ Interstate/ highway driving
- ◻ Rain
- ◻ Night-time driving
- ◻ Left hand turns against traffic
- ◻ None
- ◻ Other (list)____

8. Has the driver been involved in a crash in the past 3 years? (If you mark "No", go to question #10)

- ◻ No
- ◻ Yes

9. How many crashes was the driver involved in during the past 3 years (as a driver, not passenger)?

- ◻ 1
- ◻ 2
- ◻ 3
- ◻ 4
- ◻ 5 or more

10. How many moving violations, citations, or traffic tickets has the driver had in the past 3 years? (If you mark “0”, go to question #12)

- □ 0
- □ 1
- □ 2
- □ 3
- □ 4
- □ 5 or more

11. What moving violations, citations, or traffic tickets did the driver receive in the past three years? (Check all that apply)

- □ Failure to yield
- □ Going too slowly
- □ Not obeying traffic lights
- □ Not obeying traffic signs (such as stop sign)
- □ Improper passing
- □ Improper turning
- □ Careless driving
- □ Reckless driving
- □ Driving under influence of drugs or alcohol (DUI/DWI)
- □ Speeding
- □ Tailgating
- □ Do not know
- □ Other (list)____

12. When did the driver last attend a driver education, training, or retraining course? (If you mark “Never”, go to question #14)

- □ Within the past year
- □ 1 – 3 years ago
- □ More than 3 years ago
- □ Never
- □ Do not know

13. If they attended a driver education class, training, or re-training, what type was it? (Check all that apply)

- □ On-line class
- □ Classroom course for all drivers
- □ Classroom course for mature drivers
- □ Course with classroom and behind the wheel instruction
- □ Do not know
- □ Other (list)_______________________________________

14. How does the driver keep up with changes in road rules or laws? (Check all that apply)
    - □ Driving class
    - □ Newspaper
    - □ TV
    - □ Driver’s handbook
    - □ Friends or family
    - □ Computer
    - □ Police or law enforcement
    - □ Driver’s license office (DMV)
    - □ None of the above
    - □ Do not know
    - □ Other (list)____

BASIC DRIVING SKILLS

15. Does the driver have difficulty opening the car door?
    - □ No
    - □ Yes
16. Does the driver have difficulty getting into his or her car?
    - □ No
    - □ Yes
17. Is it difficult for the driver to adjust the car mirrors appropriately?
    - □ No
    - □ Yes
18. Does the driver have difficulty adjusting the driver’s seat so he or she can see above the steering wheel?
    - □ No
    - □ Yes
19. Is the driver able to reach the gas pedal (accelerator) and brake pedal?
    - □ No
    - □ Yes
20. Does the driver remember to turn on the headlights before driving in the dark?
    - □ No
    - □ Yes

21. Does the driver use their seatbelt?
    - ◻ Always
    - ◻ Often
    - ◻ Sometimes
    - ◻ Rarely
    - ◻ Never

22. Can the driver press the gas pedal or the brake pedal when intended?
    - ◻ No
    - ◻ Yes

23. Is it difficult for the driver to drive during daylight hours?
    - ◻ No
    - ◻ Yes

24. Does the driver place the car in the correct gear (such as drive or reverse)?
    - ◻ No
    - ◻ Yes

25. Does the driver have any difficulty turning the steering wheel?
    - ◻ No
    - ◻ Yes

26. Does the driver stop for pedestrians crossing the street?
    - ◻ No
    - ◻ Yes

27. Is it difficult for the driver to drive in good weather?
    - ◻ No
    - ◻ Yes

28. As the driver on a long trip, how frequently would the driver take breaks?
    - ◻ Every 1 to 2 hours
    - ◻ Every 3 to 4 hours
    - ◻ Every 5 to 6 hours
    - ◻ Rarely or Never

29. Does the driver have trouble staying awake while driving?
    - ◻ No
    - ◻ Yes

ALTERNATIVE TRANSPORTATION

30. Does the driver use alternative transportation (such as taking a bus or taxi)?

- ◻ Always
- ◻ Often
- ◻ Sometimes
- ◻ Rarely
- ◻ Never

31. Do you think the driver would consider alternative transportation if it were available?

- ◻ No
- ◻ Yes

FTDS Version II
6/13/2012

C: Fitness-to-Drive Screening Measure

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

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…				
A. 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. Stay in the proper lane?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
2. Check for a clear path when backing out from a driveway or parking space?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
3. Use the car controls (such as the turn signals, windshield wipers, emergency brake, or headlights)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
4. Check car mirrors when changing lanes?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
5. Read road signs far enough in advance to react (such as make a turn)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
6. Obey varied forms of traffic lights (such as green arrow for turn lane or flashing lights)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
7. Drive and hold a conversation with one or more passengers?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
8. Drive with a passenger who is providing driving directions or assistance?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
9. Drive in light rain?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
10. Drive on a highway with two or more lanes in each direction?	**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…				
11. Keep up with the flow of traffic?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
12. Keep distance from other vehicles when changing lanes?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
13. Change lanes in moderate traffic?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
14. Drive cautiously (to avoid collisions) in situations when others are driving erratically (such as speeding, road rage, crossing lane lines or driving distracted)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
15. Brake at a stop sign so car stops completely before the marked line?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
16. Maintain lane when turning (not cut corner or go wide)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
17. Back out of parking spots?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
18. Enter the flow of traffic when turning right?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
19. Share the road with vulnerable road users such as bicyclists, scooter drivers, motorcyclists?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
20. Drive on graded (unpaved) road?	**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…				
21. Check blind spots before changing lanes?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
22. Drive with surrounding tractor trailers (transport trucks)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
23. Merge onto a highway?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
24. Use a map while driving?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
25. Make a left hand turn crossing multiple lanes and entering traffic (with no lights or stop signs)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
26. Parallel park?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
27. Stay within the lane markings unless making a lane change?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
28. Stay within proper lane in the absence of road features such as clearly marked lane lines, reflectors or rumble strips?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
29. Keep distance between his or her car and others (allow time to react to hazards)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
30. Look left and right before crossing an intersection?	**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…				
31. Drive in a construction zone?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
32. Drive in dense traffic (such as rush hour)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
33. Pass (overtake) a car in the absence of a passing lane?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
34. Pass (overtake) a larger vehicle such as a RV, tractor-trailer (transport truck), or dump truck in the absence of a passing lane?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
35. Drive in an unfamiliar urban area?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
36. Control his or her car when going down a steep hill?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
37. Exit an expressway, or inter-state from a left-hand lane?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
38. Drive in a highly complex situation (such as a large city with high-speed traffic, multiple highway interchanges and several signs)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
39. Control the car (brake hard or swerve) to avoid collisions?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
40. Drive a different car (such as another person's car or a rental 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…				
41. Alter his or her driving in response to changes in health (such as vision, reaction time, fatigue, thinking, joint stiffness, medications)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
42. Drive when upset (anxious, worried, sad or angry)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
43. Stay focused on driving when there are distractions (such as radio, eating, drinking, pet in the car)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
44. Drive in an unfamiliar area?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
45. Drive at night?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
46. Avoid dangerous situations (such as car door opening, car pulling out, road debris, or an animal darting in front of car)?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
47. Drive when there is fog?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
48. Drive at night on a dark road with faded or absent lane lines?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
49. Drive when there is glare or the sun is in his or her eyes?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑

<u>FOR THE PERSON YOU ARE RATING</u>, BASED ON THE LAST 3 MONTHS, HOW DIFFICULT IS IT FOR HIM OR HER TO…				
50. Turn left across multiple lanes when there is no traffic light?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
51. Drive in a thunderstorm with heavy rains and wind?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
52. Control his or her car on a wet road?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
53. Drive on a snow covered road?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑
54. Drive on an icy road?	**Very Difficult** ❑	**Somewhat Difficult** ❑	**A Little Difficult** ❑	**Not Difficult** ❑