
Sherrilene Classen, Sandra M. Winter, Craig A. Velozo, Michel Bédard, Desiree N. Lanford, Babette Brumback, Barbara J. Lutz

OBJECTIVE. We report on item development and validity testing of a self-report older adult safe driving behaviors measure (SDBM).

METHOD. On the basis of theoretical frameworks (Precede–Proceed Model of Health Promotion, Haddon’s matrix, and Michon’s model), existing driving measures, and previous research and guided by measurement theory, we developed items capturing safe driving behavior. Item development was further informed by focus groups. We established face validity using peer reviewers and content validity using expert raters.

RESULTS. Peer review indicated acceptable face validity. Initial expert rater review yielded a scale content validity index (CVI) rating of 0.78, with 44 of 60 items rated ≥0.75. Sixteen unacceptable items (≤0.5) required major revision or deletion. The next CVI scale average was 0.84, indicating acceptable content validity.

CONCLUSION. The SDBM has relevance as a self-report to rate older drivers. Future pilot testing of the SDBM comparing results with on-road testing will define criterion validity.

In a mobile society in which older North Americans are dependent on the automobile as the primary mode of transportation, periodic assessment of their driving safety is a key concern. Currently, this assessment is taking place through comprehensive driving evaluations (CDEs) administered by driving rehabilitation specialists; by community-based programs, such as those offered by AARP, the American Automobile Association (AAA), the American Occupational Therapy Association (AOTA), and the American Society on Aging; family and caregiver assessments; and self-assessments. Self-report measures (based on self-assessment) and proxy measures (based on observations from family, caregivers, or informed lay people) are means of identifying older adults’ safe driving behaviors, increasing driving safety awareness and knowledge, and promoting behavior change and safer driving outcomes (i.e., fewer crashes, injuries, or deaths; Classen et al., 2007; Eby, Molnar, Shope, Vivoda, & Fordyce, 2003; McGee & Tuokko, 2003). Self-report measures, however, are limited by self-selection bias (i.e., capable people are more likely to complete the self-report) and social desirability bias (i.e., people are more likely to give answers that will be viewed favorably by others; Sundstrom, 2005; Zhou & Lyles, 1997). The purpose of this study was to report on item development and validity of a new self-report measure for older adults: the Safe Driving Behavior Measure (SDBM).

Safe Driving

Safe driving is the outcome of a sequence of events and interactions among events occurring at the person, vehicle, and environment levels. Safety can be...
characterized by the absence of near misses, errors, violations, crashes, and crash-related injuries or deaths. Safe driving (or lack thereof) for older adults has traditionally been measured by examining citation or violation history, crash rates, and crash-related morbidity and mortality rates (Centers for Disease Control and Prevention [CDC], 1999, 2007; National Highway Traffic and Safety Administration [NHTSA], 2008).

Understanding driving behavior sets the foundation for developing a self-report tool for older adult drivers. Driving behavior is complex and involves multiple underlying components, such as attention, cognition, decision making, and vehicle control (Michon, 1985). Older driver assessments need to capture specific driving characteristics, including patterns of driving behaviors and driving errors. For example, compared with younger drivers, older drivers (ages 65 and older) tend to engage in fewer risk-taking behaviors (e.g., speeding, driving under the influence of alcohol, or not using a seatbelt). Findings from crash data, however, indicate that older adults have some salient areas of limitations, primarily in relation to left turns against oncoming traffic and failure to yield the right-of-way at intersections (McGwin & Brown, 1999). Moreover, if involved in crashes, older drivers are at greater risk of injury than younger motorists (Awadzi, Classen, Hall, Duncan, & Garvan, 2008; Bédard, Guyatt, Stones, & Hirdes, 2002; CDC, 1999; Dellinger, Kresnow, White, & Sehgal, 2004; Zhang, Fraser, Lindsay, Clarke, & Mao, 1998).

The accepted industry standard in North America and elsewhere for assessing driving errors and determining safety, especially for older adults with health impairments, is the CDE (AOTA, 2005; Canadian Association of Occupational Therapists [CAOT], 2005; Di Stefano & Macdonald, 2005; Korner-Bitensky, Gelinas, Man-Son-Hing, & Marshall, 2005). The CDE is a battery of sensory, motor, and cognitive clinical tests and an on-road evaluation conducted by a driving rehabilitation specialist, usually an occupational therapist with specialty training. CDEs have several limitations, such as time needed to complete, high out-of-pocket costs for drivers, exposure to risks, need for trained personnel to administer the test, expensive equipment (e.g., a dual brake-equipped vehicle), liability issues, and limited opportunity for access (Kua, Korner-Bitensky, & Desrosiers, 2007; Wang & Carr, 2004). Depending on the state or province, the driving evaluator may be legally or ethically required to report unsafe performance to the licensing authority, and the driver may lose his or her license as a result. Moreover, the CDE occurs on the individual level and is therefore not a viable option to serve the safety needs of the growing numbers of aging drivers, estimated at more than 33 million in the United States and Canada (NHTSA, 2008; Transport Canada, 2007).

Self-Report

Because of these limitations, a valid self-report and proxy measure may provide an excellent opportunity for targeting older drivers who have unsafe driving behaviors and for soliciting feedback from their caregivers. A concern with using self-reports to assess driving is the issue of selection bias and social desirability bias inherent to this method of assessment. Selection bias occurs when people completing the self-report differ from those who do not (i.e., people confident about their driving may be more likely to complete a driving self-report; Zhou & Lyles, 1997). Social desirability bias affects self-reports if respondents answer according to how they would like to be perceived (typically, overstating desirable behavior and understating undesirable behavior; Fowler, 1995; Lajunen & Summala, 2003). For example, Marottoli and Richardson (1998) reported that older adults showed high confidence for driving and tended to overrate their driving ability compared with the results of a driving evaluator. Bias and error in self-reports can also come from a respondent’s poor understanding of a question, problems remembering the behavior well enough to respond accurately (recall bias), and concern about disclosing undesirable or unsafe behaviors (Fowler, 1995). These issues can be addressed by using modern statistical adjustments that account for anticipated response bias, by obtaining a proxy report in addition to self-report, and by using measures that are validated against a “gold standard” criterion measure (Snow, Cook, Lin, Morgan, & Magaziner, 2005; Sundstrom, 2008).

Valid self-report measures may yield multiple benefits, such as ease of use, maintenance of confidentiality, and immediate feedback in a safe environment (such as the home). A driver can gain insight into his or her driving safety status without risking the loss of his or her driver’s license. Most important, self-report can be made available at little or no cost to the broad older driver population in the United States and abroad (Classen et al., 2007; Eby et al., 2003; McGee & Tuokko, 2003). Moreover, from a measurement standpoint, self-reports have shown positive correlations with an on-road evaluation (Pachana & Petriwskyj, 2006; West, French, Kemp, & Elander, 1993).

Among available self-report tools, four are designed for drivers’ individual use outside of a driver improvement course or research setting. Those measures are the
Driving Decisions Workbook (Eby et al., 2003), Older Driver Skill Assessment and Resource Guide (AARP, 1992), Drivers 55 Plus: Check Your Own Performance (AAA Foundation for Traffic Safety, 1994), and the computer-based Roadwise Review (AAA, 2004). Strengths of the measures include assessment of driving that includes medical conditions and medication use (Driving Decisions Workbook), measurement of physical and cognitive abilities predictive of at-fault crashes (Roadwise Review), and education for drivers about risk factors and strategies for driving safely (Driving Decisions Workbook, Drivers 55 Plus: Check Your Own Performance, Older Driver Skill Assessment and Resource Guide, and Roadwise Review). Only the Driving Decisions Workbook development entailed comparison with a criterion measure to determine the influence of self-report bias (e.g., social desirability; Eby et al., 2003). The Driving Decisions Workbook and Roadwise Review both incorporated focus group feedback from stakeholders during measure development (Eby et al., 2003; Staplin & Dinh-Zarr, 2006). AARP’s measure provides a score outcome as part of feedback for drivers; it has been criticized, however, for a lack of published psychometric testing in establishing the scoring system’s reliability and validity (Eby, Trombley, Molnar, & Shope, 1998).

Limitations of paper measures designed with an educational focus (i.e., emphasizing safety knowledge such as driving strategies) include their long length (20–47 pages), which increases respondent burden (AAA Foundation for Traffic Safety, 1994; AARP, 1992; Eby et al., 2003). The computer-based Roadwise Review takes approximately 40 min to complete, may be challenging for older adults with low computer fluency to use, and requires the assistance of another person to complete (Staplin & Dinh-Zarr, 2006). Last, none of these measures was developed from a socioecological framework; the current measures have a person focus and lack adequate vehicle and environment items.

Self-report challenges include making the tool available for the target population, designing a tool that supports self-administration, and ensuring that the behavioral measures are relevant for a broad group of older drivers. A need emerges for a valid and reliable self-report and proxy report to identify safe and unsafe driving behaviors in older adults, but self-reports must provide feedback on strategies for driving safer and longer or indicate the necessary referrals to address problems. This need can be more comprehensively fulfilled by using community-based participatory research (including stakeholders’ opinions), using traditional measurement methods (building a valid and reliable instrument), and using Item–Response Theory (IRT; providing meaningful and measureable descriptions at the level of the person and the item). The contribution of community-based participatory research (CBPR), traditional measurement methods, and IRT are discussed next.

Community-Based Participatory Research

The CBPR research approach is based on acknowledging the community, tapping its expertise and knowledge, and partnering with the community in the creation and implementation of research to address a public health problem (Israel, Schulz, Parker, & Becker, 1998). Potential contributions of participants in developing a measure include identifying key behaviors and providing feedback on the understandability of items (Stone, 1997; Streiner & Norman, 2003). The CBPR approach captures the real-life phenomena under study by using the knowledge and perceptions of people with first-hand experience—in this case, people such as older drivers, families, and other community members with an interest in the driving safety of older adults (Minkler & Wallerstein, 2003). In developing a driving self-report, description of the real-life phenomena by community members can help explain driving differences across settings related to geographic variations (e.g., terrain or weather) and cultural norms. During development of a measure, processes that engage community participants, such as focus groups and feedback from an advisory committee, can be integrated with measurement theory principles to improve the acceptability, relevance, clarity, understandability, cultural relevance, and formatting of the items and measure—all of which enhance its validity (Vogt, King, & King, 2004).

Measurement Theory

Items developed for a measure must have acceptable face and content validity. Face validity is an initial judgment of whether a tool assesses the behaviors it is supposed to (Streiner & Norman, 2003). Carmines and Zeller (1979) defined a measure with content validity as one whose collective items adequately represent the construct under investigation. Defining the construct and related domains for inclusion in a measure using theoretical frameworks is a critical step in establishing face and content validity (Lynn, 1986). After development of an initial item set, judgment of the items by outside raters is essential for further item revisions and improvements. Content validity is assessed by a quantification of item and measure relevance obtained from expert raters using a content validity index (CVI; Lynn, 1986).
Although validity testing can improve a tool’s utility, acceptability, and item relevance, traditional methods have limitations when the goal is development of accurate items to precisely and objectively estimate a person’s function. For example, a precise and objective measurement of driving function requires assessing the difficulty a person has with driving situations of differing complexity, such as parking and merging. Therefore, in addition to traditional measurement methods addressing validity, use of an item-response theory (IRT) approach offers advantages (Bond & Fox, 2007).

IRT is a measurement model that relates item difficulty for a measure to person ability, thereby enabling more precise measurement (Bond & Fox, 2007). IRT has been used in driving studies to develop driving scales and to analyze behaviors observed in on-road testing (Kay, Bundy, Clemson, & Jolly, 2008; Myers, Paradis, & Blanchard, 2008). In constructing a measure using IRT, behaviors representing the construct under consideration (e.g., safe driving) are outlined on a continuum of item difficulty for people with a greater or lesser capacity (Stone, 1997). Items are designed with a range and specificity to maximally separate people on the basis of ability. For example, we propose that people who are not safe drivers would have difficulty with relatively “easy” items such as staying in their designated lane when driving. Conversely, people who are safer drivers will consistently accomplish easy items and may only report difficulty with “hard” items such as controlling their car on a snowy road. Therefore, given such benefits of IRT as obtaining increased precision, the IRT method becomes particularly useful when measuring a functional behavior such as driving.

Rationale and Significance

No valid and reliable self-report or proxy measure of older driver safety behaviors exists that is accessible, relevant, culturally sensitive, geographically representative, and appropriate for a broad population of older drivers in North America (the long-term focus of this project). Clearly, such a safe driving behavior measure has utility for older drivers, concerned family members, caregivers, aging advocates, and professionals on a communitywide and population-wide level.

Purpose

Using traditional measurement methods and IRT, as well as CBPR principles, our objective is to report on item development and validity of a new self-report measure for older adults: the SDBM. Use of this measure will position occupational therapists and agents of the aging networks to influence older driver safety at the population level.

Method

Setting

Two geographically diverse and culturally rich sites (in Florida and Ontario) participated in this study. The participating universities, the University of Florida and Lakehead University, both have a driving evaluation and research center with an established record of older driver research.

Research Team

The research team consisted of five doctoral-level researchers with combined research expertise in older driver safety, public health, measurement theory, item development, CBPR, and biostatistics (Classen, Velozo, Bédard, Brumback, and Lutz). Two additional team members participated in the development, refinement, and testing of the SDBM: one an occupational therapist and certified driving rehabilitation specialist (Lanford), and the other an occupational therapist and doctoral candidate in rehabilitation science (Winter). For this phase of the study, we included community participation from four expert raters representing the United States, Canada, and Australia and feedback from five advisory panel members representing CAOT; AARP; Transport Canada; the University of Michigan Transportation Research Institute; and AOTA’s Older Driver Initiative, which is funded in part by NHTSA. To build the item bank, we also solicited the opinions of 14 focus group participants in Thunder Bay, Ontario, and 18 focus group participants in Gainesville, Florida.

Process for Item Development

Our methods for item development included the CBPR approach; measurement principles; guiding frameworks, including three theoretical models; previous research; published driving self-report measures; published driving evaluations from the older driver safety literature; and expert opinion. Each method is discussed in the following sections.

Precede–Proceed Model of Health Promotion. This team has an established publication record of research on older driver safety issues using the Precede–Proceed Model of Health Promotion (PPMHP; Green & Kreuter, 2005). The PPMHP domains address safety influences on the level of the driver (e.g., knowledge, behavior, attitudes, health status; Awadzi et al., 2008; Classen et al., 2006, 2007), the vehicle (e.g., safety features or equipment;
Awadzi et al., 2008; Classen et al., 2007), the physical and political environment (e.g., daytime conditions, state licensing policies; Awadzi et al., 2008; Classen et al., 2007), the social environment (e.g., driving laws, assistance with transportation; Classen, Winter, & Lopez, 2009), and person–vehicle–environment interactions (Classen, Awadzi, & Mkanta, 2008). Collectively, our research has demonstrated that behavioral and environmental factors are understudied yet important determinants for older driver safety and that a comprehensive assessment must include person, vehicle, and environmental factors, as well as the interactions among these factors (Classen et al., 2007, 2008).

**Haddon’s Matrix.** Haddon’s (1972) Matrix assumes that injury is a preventable occurrence and provides a framework to target risk or protective factors that can occur at the person level, means of injury transmission level, environment level (physical and social driving context), and vector level (physical forces). Haddon conceptualized behaviors related to injury on a time continuum. For the purpose of this study, we defined a time continuum that focused on behaviors at two points: before driving (e.g., checking vehicle) and during driving (e.g., driving maneuvers).

**Michon’s Model.** Michon’s (1985) model of driving behavior established a hierarchy of driving behavior on three levels: strategic, tactical, and operational. Strategic behaviors can occur before or during driving, include general goal formation decisions (e.g., trip planning), and can occur in minutes to hours. Tactical behaviors reflect conscious decisions made while operating the vehicle relating to interaction with the immediate driving environment (e.g., maneuvers or car handling) and can occur in seconds to minutes. Operational behaviors are outputs that happen primarily by subconscious action (e.g., maintaining your lane) and occur in seconds.

**Existing Measures.** We reviewed several driving self-report and evaluation measures (Hunt, personal communication, April 30, 2007; Justiss, 2006; Kay et al., 2008; Myers et al., 2008; Owsley, Stalvey, Wells, & Sloane, 1999; Reason, Manstead, Stradling, Baxter, & Campbell, 1990; Stalvey & Owsley, 2000; Yee & Melichar, 1992). We extracted and coded items from those measures according to our theoretical frameworks (PPMHP, Haddon’s Matrix, Michon’s model). On the basis of the PPMHP and Haddon’s Matrix, items were coded as addressing person factors (skills, knowledge, attitudes, or health status), vehicle factors (selection and use of car features), or environment factors (physical environment [e.g., weather], social environment [e.g., passengers in car]). We coded items according to when the behavior occurred (predriving or driving) and the level of the behavior per Michon’s (1985) hierarchy (strategic, tactical, or operational). We found that the items on the existing measures were biased toward person factors and provided minimal representation of environment and vehicle factors or person–vehicle–environment interactions. Moreover, most items assessed driving deficits (e.g., failure to maintain proper speed for conditions) and therefore did not represent the broad range of driving behaviors demonstrated by older adults.

**SDBM Tool Design.** We designed the SDBM as a self-report tool for use by older adults in assessing driving behavior. Modified versions of the tool were created for use by family members, caregivers, and professionals. The current paper-and-pencil version of the SDBM is an older driver self-report measure, including instructions, demographic profile (e.g., age, gender), driving history profile (e.g., driving habits and patterns), and 68 questions on driving ability. For each driving behavior item, the respondent was asked to rate his or her driving difficulty with that behavior over the previous 3 mo. To minimize recall bias (e.g., failure to remember previous driving behavior accurately), we selected 3 mo as the response time period (Fowler, 1995). For most respondents, a 3-mo driving period would incorporate a variety of driving situations while allowing for easy recall of driving behavior and difficulties encountered. We chose to focus on the respondent’s difficulty with driving behaviors (rather than frequency) because we considered difficulty to be more informative of safety (than frequency) and congruent with IRT principles. On the basis of team judgment and peer review feedback, we used five response levels for the driving behavior items: cannot do, very difficult, somewhat difficult, a little difficult, and not difficult, with a not applicable response added for items that not all drivers would encounter (e.g., control your car on a snow-covered road).

**Item Development.** Working from our earlier definition of driving behavior, team expertise, previous research findings, and existing measures, we generated item ideas representing a continuum of driving behavior from least difficult (e.g., use car controls such as headlights or turn signals) to most difficult (e.g., “control your car on an icy road”). Figure 1 outlines the step-by-step process for item development. Team members worked consistently (2–3 hr per week × 6 mo) and collaboratively to initiate, discuss, refine, reward, add, and delete ideas for items and to determine item placement along the continuum. We developed driving behavior items to provide adequate representation of person–vehicle, person–environment, and person–vehicle–environment interactions (e.g.,
“control the car to avoid collisions”) occurring across two time frames (before and during driving) and including strategic, tactical, and operational concepts.

**Item Review.** The item review process was completed in two stages. First, a group of peer reviewers (five doctoral students and four doctoral-level trained qualitative researchers from the Qualitative Data Analysis Group, College of Nursing, University of Florida) provided input and recommendations on the face validity of the items and the utility of the measure (e.g., ease of completing items and time needed). Following the guidelines of Lynn (1986), who suggested that three or more raters are adequate to provide a rigorous rating, we asked four expert raters from our advisory committee to complete a CVI. They rated the relevance of each SDBM item on a 4-point Likert scale (1 = not relevant, 2 = relevant with major revisions, 3 = relevant with minor revisions, and 4 = very relevant) and gave feedback on item accuracy, purpose, organization, clarity, appearance, understandability, and adequacy (Grant & Davis, 1997).

**Analysis**

Our analysis plan included calculation of both an item-level CVI and the scale average CVI. Using CVI procedures (Lynn, 1986), rater scores were collapsed with an item-level score of 3 or 4, indicating acceptable item relevance, and a score of 1 or 2, indicating need for a major revision or low item relevance. After entering all ratings and comments in a spreadsheet, we calculated item-level CVI scores (i.e., the proportion of four raters who scored the item as relevant, a rating of 3 or 4) and the scale average CVI. Item-level CVIs of 0.75 or 1.00 were acceptable (0.75 = the item was rated as relevant by three raters; 1.00 = the item was rated as relevant by four raters), whereas scores of 0.5, 0.25, or 0.0 (0.5 = the item was rated as relevant by two raters, 0.25 = the item was rated as relevant by one rater, and 0.0 = no raters indicated relevance) were unacceptable. We used the statistical package R: A Language and Environment for Statistical Computing, version 2.5.0 (R Development Core Team, 2008) for computing the CVI. After the analysis, items with a low item CVI (≤0.5) were revised by the research team. The CVI process and item refinement were repeated until an acceptable level of content validity was reached (average CVI ≥ 0.80; House, House, & Campbell, 1981).

**Results**

**Peer Review**

The peer review feedback indicated favorable face validity and utility while suggesting revisions for the format of the tool and 11 (18%) of the 60 SDBM items. Specific peer review results pertained to clarity of instructions, wording, relevance, or response level of items.

**Content Validity Index**

First-round CVI results indicated 44 (73%) of the 60 SDBM items had an acceptable item-level CVI ≥ 0.75 (i.e., a majority of the four raters had rated the item as relevant). Sixteen (27%) of the items had an item-level CVI of ≤0.5, indicating either a lack of relevance or the need for a major revision of that item. The scale average CVI was 0.78, falling short of the acceptable scale average of 0.80. Collectively, the item-level CVIs and the scale average CVI indicated a need for revisions and a repeat of the expert review and CVI. We therefore revised the 16 items with a CVI rating of ≤0.5 as follows: Six items with low relevance were deleted, one item was moved to Section A (demographic questions), one item was grouped in the scale with items of similar difficulty, six items were reworded, and two items were retained without revision on the basis of team judgment and clinical or...
Table 1. Items and Content Validity Index (CVI) Before (Initial) and After (Final) Team Revisions and Expert Ratings

<table>
<thead>
<tr>
<th>Item</th>
<th>Initial CVI</th>
<th>Item revision</th>
<th>Final CVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasten your seatbelt</td>
<td>0.5</td>
<td>Moved item to Section B: Driving history profile</td>
<td>Not rated</td>
</tr>
<tr>
<td>Check mirrors when backing out from driveway or parking space</td>
<td>0.5</td>
<td>Revised wording to incorporate both check mirror AND turn head</td>
<td>1.0</td>
</tr>
<tr>
<td>Locate the control for your hazard lights</td>
<td>0.5</td>
<td>Deleted in concurrence with rater feedback of low relevance</td>
<td>Not rated</td>
</tr>
<tr>
<td>Keep both hands on the wheel while driving</td>
<td>0.5</td>
<td>Deleted in concurrence with rater feedback of low relevance</td>
<td>Not rated</td>
</tr>
<tr>
<td>Pull out smoothly from a stop sign with no harsh acceleration</td>
<td>0.25</td>
<td>Deleted in concurrence with rater feedback of low relevance</td>
<td>Not rated</td>
</tr>
<tr>
<td>Maintain car speed within 5 mi/8 km per hr over or under the posted speed</td>
<td>0.5</td>
<td>Reworded to follow CVI rater suggestion of “Back out of parking spots”</td>
<td>0.75</td>
</tr>
<tr>
<td>Drive with tractor trailers (transport trucks) around you</td>
<td>0.5</td>
<td>Deleted because the behavior is covered by another SDBM item “Keep up with flow of traffic”</td>
<td>Not rated</td>
</tr>
<tr>
<td>Scan left to right at intersection</td>
<td>0.5</td>
<td>Retained on the basis of team judgment</td>
<td>0.5</td>
</tr>
<tr>
<td>Complete a 100-mi (160-km) drive on your own</td>
<td>0.5</td>
<td>Reworded to “Look left and right before crossing an intersection”</td>
<td>1.0</td>
</tr>
<tr>
<td>Drive up or down a steep hill</td>
<td>0.5</td>
<td>Deleted in concurrence with rater feedback of low relevance</td>
<td>Not rated</td>
</tr>
<tr>
<td>Drive when you are upset</td>
<td>0.5</td>
<td>Reworded to clarify the term upset as anxious, worried, sad, or angry</td>
<td>0.75</td>
</tr>
<tr>
<td>Drive after taking medication</td>
<td>0.5</td>
<td>Moved item to Section B: Driving history profile</td>
<td>Not rated</td>
</tr>
<tr>
<td>Drive in an unfamiliar area</td>
<td>0.5</td>
<td>Retained on the basis of team judgment</td>
<td>0.5</td>
</tr>
<tr>
<td>Drive through a railroad crossing</td>
<td>0.5</td>
<td>Deleted in concurrence with rater feedback of low relevance</td>
<td>Not rated</td>
</tr>
</tbody>
</table>

We obtained a second round of ratings for the CVI to reassess 5 revised items and to obtain initial ratings of 21 new items added after three focus groups conducted at the Florida and Ontario sites. Second-round CVI results indicated that the 5 revised SDBM items had an acceptable item-level CVI $\geq 0.75$. For the 21 new items, 15 had an acceptable item-level CVI $\geq 0.75$, whereas 6 of the items had an item-level CVI of $\leq 0.5$, indicating either a lack of relevance or the need for a major revision of that item. Therefore, we revised the 6 items with a CVI rating of $\leq 0.5$ as follows: 1 item with low relevance was deleted, 2 items were moved to Section B: Driving history profile, and 3 items were reordered after rater feedback. After these revisions, the final scale average CVI on the retained items from the first- and second-round ratings was 0.84. The final scale average was above the acceptable scale average level of 0.80 (House et al., 1981), indicating relevance of the SDBM items.

The item pool is currently distributed across three main domains: Section A–Demographic Profile, Section B–Driving History Profile, and Section C–SDBM.1 The item pool in Section C yields a combination of items capturing behaviors across the span of person, vehicle, and environment. For example, Section C captures 68 safe driving behaviors classified in the person–vehicle domain (11 items), person–environment domain (42 items), and person–vehicle–environment domain (15 items).

Discussion

The population of older drivers who would benefit from a self-report measure designed to provide targeted feedback on their driving ability is growing. The purpose of this study was to describe the item development of and to report on validity analyses and rater agreement for the SDBM, a self-report measure for older drivers. This study illustrates the use of theoretical frameworks for developing a self-report for measuring safe driving behavior and indicates limitations of existing self-report measures in meeting the need for precise measurement tied to driving ability. The SDBM is unique among self-report and proxy measures in that we used a combined measurement approach with traditional measurement strategies addressing validity and IRT strategies to increase the precision of items. Moreover, we used a CBPR approach to inform item development and contribute to geographic and cultural relevance of the items.

Development of the SDBM was informed by feedback from a peer review group and two rounds of expert rater review. In describing the process for development of the SDBM, we intend to further the body of knowledge on self-report and proxy measures for use in promoting awareness of safe driving behavior by older adults or their caregivers. The face and content validity of this tool, as

---

1The final version of the SDBM is still undergoing psychometric testing, but a copy of the current measure is on file and available from Sherrilene Classen: sclassen@phhp.ufl.edu.
well as the relevance determined by expert raters during two rounds of reviews, provide early support for the utility of this item pool and for further developing the measure.

Limitations of self-report will be accounted for in the three following ways. First, we are implementing steps to minimize the effects of selection bias by providing a tool that can be used privately and confidentially. Second, we will manage social desirability bias by comparing results with a companion proxy report from a caregiver and by conducting formal on-road driving assessments from which we will discern criterion validity between the self-report, proxy report, and on-road test. Third, we are managing recall bias by asking participants about their driving behaviors in the past 3 mo. Finally, we conducted rater reviews only on the revised items (from previous rater feedback) and new items (from focus groups) during the second round of reviews but plan future psychometric evaluation on the full item set.

Strengths of the SDBM include theoretical support from three frameworks (Precede–Proceed, Haddon’s Matrix, and Michon’s Model), a collaborative item development process that solicited feedback with revisions completed at several time points, and a final rater agreement of 84% on the validity of the tool. Finally, the items have been developed to capture behaviors across the span of the person, vehicle, environment, and their interactions.

With future research, including testing in a multisite study, we aim to provide a valid and reliable tool informing older adults, family members, caregivers, and professionals about driving behavior and safety. We assert that given accurate knowledge about an older adult’s level of driving safety, safety actions such as referrals for driving evaluation or remediation will be more readily accomplished.

Acknowledgment
The project is funded by a National Institutes of Health/National Institute on Aging grant “Community Participation in Developing a Measure for Safe Older Driver Behaviors.” We acknowledge the National Older Driver Research and Training Center at the University of Florida and the Centre for Research on Safe Driving at Lakehead University, Ontario, Canada. We wish to thank people who assisted with this study, including the expert raters: David Eby, Bella Dinh-Zarr, Holly Tuokko, and Jim Langford, and peer reviewers from the Qualitative Data Analysis Group at the University of Florida led by Barbara Lutz and Sharleen Simpson. We also wish to recognize our advisory board members: Paul Boase, Jan Polgar, David Eby, Barbara Messinger-Rapport, and Frank Carroll. Co-author Sandra Winter completed this project in partial fulfillment of a PhD degree in rehabilitation science from the University of Florida.

References
Classen, S., Winter, S., & Lopez, E. D. S. (2009). Meta-synthesis of qualitative studies on older driver safety and
Grant, J. S., & Davis, L. L. (1997). Selection and use of
Dellinger, A. M., Kresnow, M. J., White, D. D., & Sehgal, M.
Fowler, F., Jr. (1995). Improving survey questions: Design and
Eby, D. W., Trombley, D. A., Molnar, L. J., & Shope, J. T.
Haddon, W. J., Jr. (1972). A logical framework for categoriz-
Korner-Bitensky, N., Gelinas, I., Man-Son-Hing, M., &


