EVALUABILITY ASSESSMENT OF A NATIONAL DRIVER RETRAINING PROGRAM: ARE WE EVALUATING IN THE RIGHT LANE?

Mélanie Joanisse  
Arne Stinchcombe  
Stephanie Yamin  
University of Ottawa  
Ottawa, Ontario

Abstract: An evaluability assessment (EA) of the 55 Alive program, a national older driver refresher course aimed at improving driving skills, was conducted. This EA adds to the evaluation literature as previous outcome evaluations neglected to explore whether this program was prepared for such assessments. A mixed-method protocol was executed across three community sites. Based on the results of this EA, several suggestions for future evaluations are discussed: (a) sampling all stakeholder levels for a more holistic snapshot of the program, (b) using EA to facilitate stakeholders’ engagement in the evaluation, and (c) incorporating scientists specializing in the area within the evaluation.

Résumé : Une évaluation de l’évaluabilité (EE) du programme 55 ans au Volant, un programme visant à améliorer les habiletés des conducteurs âgés, a été entreprise. Cette EE s’ajoute à la littérature dans le domaine car des évaluations antérieures de résultats ont négligé d’investiguer si ce programme se prêtait à ce type d’évaluation. Un protocole requérant des méthodes mixtes a été utilisé auprès de trois sites communautaires. Des recommandations pour les évaluations futures découlent des résultats de cette EE : (a) l’importance d’un échantillonnage comprenant tous les niveaux des parties prenantes, (b) l’utilisation de l’EE afin de faciliter l’engagement des parties prenantes, et (c) l’inclusion de chercheurs se spécialisant dans le domaine évalué.

Older adults constitute the fastest growing segment of the Canadian population and, as the “baby boomers” age, the Canadian senior population is expected to reach 6.7 million in 2021 (Health Canada: Division of Aging and Seniors, 2002). This increase will inevitably translate into a greater number of older drivers. In Ontario, for example, it is estimated that the number of older drivers will
reach approximately 2.5 million by 2028 (Hopkins, Kilik, Day, Rows, & Tseng, 2004). This is an important trend to consider as drivers aged 65 years and older have the greatest risk for motor vehicle collisions (MVC), once distance driven is considered (Barr, 1991), and the highest risk of driving-related fatalities per mile driven (Dulisse, 1997). Driving is associated with a sense of autonomy and independence, which contribute to a sense of well-being, an ability to preserve social contacts, and contribute to overall quality of life (Ragland, Satariano, & MacLeod, 2004). It follows that efforts to maintain safe driving abilities among older adults are necessary. To this end, one remedial strategy to decrease the frequency of MVCs may be driver education that focuses on the needs of older drivers.

At first glance, the logic behind older driver retraining might seem rather obvious: providing instruction aimed at improving driving abilities would lead to older drivers becoming safer road users. A number of programs following this model have been developed, ranging from physical exercise and on-road training to psycho-educational courses such as the 55 Alive course (Eby, Molnar, Shope, Vivoda, & Fordyce, 2003; Janke, 1994; Owsley, McGwin, Phillips, McNeal, & Stalvey, 2004; Owsley, Stalvey, & Phillips, 2003; Ostrow, Shafron, & McPherson, 1992; Roenker, Cissell, Ball, Wadley, & Edwards, 2003). Unfortunately, the intuitive appeal of such programs is insufficient to justify their implementation, especially on a widespread basis. Rigorous experimentation and continual efforts to improve the program based on empirically derived evidence are required, all of which may be accomplished through program evaluation. To this end, the objective of this paper is to report the findings from an evaluability assessment (EA) of the 55 Alive mature driver-refresher course offered in Canada by the Canada Safety Council (CSC).

HISTORICAL CONTEXT

The American Association for Retired Persons (AARP) created the 55 Alive program in order to assist older drivers to improve their skills and ultimately prevent traffic accidents. To date, over 13 million students have graduated from this program (AARP, 2004). In Canada, the need for a course that focuses specifically on older drivers’ difficulties was recognized at a national colloquium comprised of professionals, older adults from the community, and the Canada Safety Council (CSC). Specifically, older drivers are particularly interested in a driving refresher course, as some do not receive formal driving training.
and most provinces require a written drivers’ test once drivers reach the age of 80 years. Maintenance of a drivers’ licence is of crucial importance given the negative outcomes linked to driving cessation. The CSC, a non-governmental, non-profit organization with an official mandate to prevent deaths and injuries at work, at home, and in traffic, adapted AARP’s 55 Alive course for the Canadian population by modifying the content and translating the text into French (CSC, 1989). It is important to note that this article describes the evaluation of the Canadian version of the 55 Alive program.

**PROGRAM PROFILE**

The 55 Alive course is a group-based refresher course for mature drivers (i.e., all Canadian drivers aged 50 years and above) offered in both French and English. The objectives of the program as stated in the publicity material and by communications with the CSC are to increase: (a) mobility and autonomy for the older driver, (b) driving confidence, (c) knowledge and/or awareness of different driving-related topics, (d) anticipation of other drivers’ actions, and (e) discussion of driving-related concerns. The program also aims at correcting bad driving habits and improving overall driving skills. It is offered in the community for a minimal fee. The recommended class size is between 10-12 students (CSC, 2001). Students are provided with a student manual. A trained individual, who will be referred to as a facilitator, teaches the course in community centres with the use of an instructor’s manual and multimedia equipment. A total of eight topics are covered in six hours spread over two days.

The program involves psycho-education on the following topics: (a) description of mature drivers and self-assessment; (b) physical (i.e., vision and hearing) changes and how they affect driving performance; (c) interacting with traffic in normal driving situations; (d) hazardous driving environments; (e) driver guidance, using road signs and signalizations; (f) the vehicle environment; (g) alcohol and medications and their interaction with driving; and (h) driving decisions (planning for driving cessation). Once the course is completed, each student is given a wallet card, which can be used as proof that they have completed a driver-refresher program (CSC, 2001). Finally, while CSC offers the class, program managers in the community who are not employed by CSC are usually responsible for the recruitment and delivery of the program. The logic model of the program is available in the original full report of the evaluation (this report is available by contacting the authors).
PREVIOUS EVALUATIONS

To our knowledge, five evaluations of the 55 Alive driver refresher course offered by CSC have been conducted and published in peer-reviewed forums (Bédard et al., 2008; Bédard, Isherwood, Moore, Gibbons, & Lindstrom, 2004; Nasvadi, 2007; Nasvadi & Vavrik, 2007; Porter, 2007). Using an experimental design, Bédard and colleagues (2004) found that there was no improvement in participants’ driving ability (i.e., as determined by an on-road examination) as a result of completing the 55 Alive driver refresher course. The authors examined the effectiveness of the 55 Alive program in improving errors on a driving evaluation in comparison to a video feedback group and a control group. There was a significant reduction in the number of errors in the video-training group only. Nevertheless, Bédard and colleagues’ (2008) evaluation indicated that participation in the 55 Alive program might be useful in increasing participants’ knowledge and that the combination of the 55 Alive program with an on-road training component yields improvements in some elements of driving behavior (Bédard et al., 2008). These results are also supported by Porter (2007), who found no significant improvement in driving performance solely after completing the 55 Alive program. Nasvadi (2007) used a retrospective cohort design with 367 drivers between the ages of 55 and 94 years who had completed the 55 Alive program. The results suggested that the program may be useful in promoting awareness of issues pertinent to the older driver. Moreover, a follow-up study conducted by Nasvadi and Vavrik (2007) found that after completion of the program there was no significant change in crash-rate among younger mature drivers (i.e., between the ages of 55 and 75). Interestingly, a negative effect (i.e., increased crash frequency) of taking the course was observed for male course completers over the age of 75.

Despite the relevance of these findings, a number of issues remain based on the methodological limitations of the studies in question. Specifically, the internal validity of these evaluations might have been influenced by a self-selection bias. Furthermore, the retrospective nature of many of these evaluations makes it difficult to fully acknowledge the effectiveness of the findings in the absence of a control group. Finally, due to the addition of the on-road training component in Bédard and colleagues (2008), it is difficult to determine whether improvements in on-road behaviour were due to the 55 Alive program alone, the on-road component alone, or both in combination. When considered together with Bédard and
colleagues’ (2004) findings, these results seem to imply that it is the addition of the on-road training component that led to improvements in driving performance. Furthermore, it is curious why participants in the control group did not complete a paper-and-pencil-based test on driving safety. Consequently, although the article seems to suggest that the 55 Alive refresher course leads to greater knowledge of safe driving, this finding may be the result of simply re-testing individuals in the exact same manner. Finally, examination of the on-road results between sites reveals that findings were not replicated across sites, suggesting that there might have been a lack of standardized program implementation across sites. Based on this observation, it is reasonable to hypothesize that program facilitators might play a crucial role in the program’s implementation and, therefore, success.

Regardless of these methodological shortcomings, the evaluations presented are important attempts to determine the impact of the 55 Alive course. Nevertheless, the main issue might be that the evaluations took a summative approach in that they primarily focused on behavioural outcomes such as safe driving. They sought to determine the efficacy and effectiveness of the program rather than investigate the design and implementation of the program itself in order to make suggestions for improvement. This seems problematic, given that the findings suggest that the program is not being administered in the same manner across sites and that this may be affecting program outcomes. It was also found that the objectives mentioned by CSC differed from those cited in the literature, and it remains unclear that the program theory (i.e., that a psychoeducation course can improve driving performance and therefore reduce crashes) is justified.

Considering Trochim’s (2005) phased approach to program development and evaluation, one could argue that the previous evaluations of the 55 Alive driver refresher course might have “jumped the gun.” Specifically, Trochim’s approach emphasizes the importance of conducting sufficient Phase 1 (formative evaluation) research, before conducting Phase 2 (case control, quasiexperimental or correlative methods) or phase 3 (randomized) studies. We were unable to find an evaluability or formative assessment of the 55 Alive program in the literature (Phase 1). All prior evaluations have focused on outcomes and the overall impact of the program. They have failed to examine the program theory and process and whether the program is, in fact, prepared for such an evaluation (Phases 2 and 3).
CHOICE OF EVALUATION

Consequently, the present evaluation is an EA. The concept of EA mainly appeared in the 1970s when scholars noticed that some outcome evaluations of federal programs in the United States were conducted prematurely, negatively impacting their usefulness and utility. It became apparent that there was a need for an *a priori* evaluation that would help determine if a program was “evaluable” before resources were to be allocated to implementation and outcome evaluation (Evaluating Socio Economic Development, 2009; Smith, 1989; Wholey, 1979). A variety of goals and components of EA were proposed, which led to a number of separate conceptions of EA. The present evaluation drew from several sources to generate a working definition of EA. In particular, Wholey’s (1979) conceptualization of EA was selected as it emphasizes sound program theory as well as agreed-upon objectives. Second, Rossi, Lipsey, and Freeman’s (2004) conceptualization of EA was also chosen as it emphasizes the plausibility of intended goals and outcomes as well as how to operationalize those outcomes. We also included Smith’s (1989) position because it enhances this EA as it emphasizes improving the program as well as determining program readiness for an impact evaluation. In brief, this working definition highlights the comprehensiveness and flexibility of the EA framework.

EVALUATION QUESTIONS

In employing an EA, the evaluator is allowed considerable flexibility with regard to the types of evaluative questions and elements on which to focus. Based on the description of the program as well as previous literature, a number of themes have been assembled to further structure the current EA. Specifically, issues relating to the theory, process and implementation, and perceived impact of the program were evaluated. We sought to investigate whether the objectives and goals of the 55 Alive program were: (a) clear to stakeholders, (b) well-defined, and (c) plausible. We also evaluated: (d) whether the 55 Alive program served the population for whom it was designed, (e) whether the activities were being implemented by the community stakeholders as designed by CSC, (f) whether stakeholders were satisfied with the 55 Alive program, and (g) whether the program had the capacity to provide data for an outcome evaluation.
METHOD

Selecting an EA Framework and Guidelines

As noted in Trevisan (2007), it is important that evaluators indicate the steps offered by a particular EA framework. Therefore, the methodology for this EA followed Smith’s (1989) framework. Smith’s (1989) 10-step model, as summarized in Trevisan and Huang (2003), was selected, based on its clarity and its applicability in a variety of contexts. The EA followed the ethical guidelines of the Canadian Evaluation Society (2007), the guiding principles for evaluators by the American Evaluation Association (2004), and the Program Evaluation Standards from the Joint Committee on Standards for Educational Evaluation (1994). These standards assert that reliable program evaluations should be useful, feasible, proper, and accurate (Sanders, 1999).

The methodological approach chosen was a mixed-method design as it included both qualitative and quantitative approaches. However, the measures used for this EA were mainly qualitative in nature. They reflect common methods used for EA and include interviews, focus groups, observations, and questionnaires (Trevisan, 2007). In order to maximize the accuracy of the data collected, the evaluation attempted to triangulate, that is, to utilize more than two sources of information or data collection strategies. All interviews and focus groups in this EA followed the guidelines of Holcomb (1993, as cited in Wholey, Hatry, & Newcombe, 2004; full protocols are available by contacting the authors). Participants were aware that their responses would remain confidential and that the evaluators were not associated with regulating authorities or insurance companies. The interviews and focus groups were conducted informally. This evaluation received approval from the Research Ethics Board at the University of Ottawa. The authors of this article evaluated the program, therefore “evaluators” and “authors” are used interchangeably in this text.

Participants and Data Collection

CSC level

The program manager at the Canada Safety Council directly responsible for the 55 Alive program participated in this study.
Community site level

Three community sites (two English and one French) from a medium-sized city in Ontario were selected. The two English-speaking sites were selected randomly from the community sites offering the course within the specified region. The French-speaking site was chosen because it was the sole site offering the French version of the program in the specified region. Two sites offered the course during the evaluation period, while one site previously offered the course.

For the sites offering the course during the evaluation period, a focus group was conducted with the participants. Participants were queried on the theory, process, and implementation (e.g., behavioural change noticed as a result of this course and obstacles to participation). Participants completed a demographic questionnaire asking questions related to age, gender, ethnicity, education, socioeconomic status, as well as physical and mental health. Questions pertaining to driving history were also asked, such as possession of a valid driver’s license and kilometres driven annually. A total of 17 participants (Mean age = 74.24 years, SD = 9.12) were included in this EA. The evaluators observed the entire course sessions offered by the two community sites; one site offered the course in one day while another site offered the course in two half days. These observations were recorded by hand and later coded in an implementation grid developed by the evaluators. The implementation grid was designed using the guidelines outlined by the Communities That Care (2007) initiative. Specifically, the grid included elements of successful, high-fidelity implementation, such as reaching the target audience (i.e., individuals aged 50 and above) and ensuring that components of the program are delivered according to the original design (i.e, coverage of all topics) as well as at the intended dosage (i.e., duration of the program and number of sessions). Assignment of homework, participation of older adults, and use of supporting materials (e.g., use of overheads, reference to Ontario Driver’s Handbook) were also noted.

As for the site that had already given the course, phone interviews (that also contained the demographic questionnaire) were completed with five participants (M = 78.8, SD = 3.03) who had taken the course prior to our evaluation and whose contact information was provided to us by the community sites.

Two of the program managers of the selected community sites were interviewed by phone, while one program manager responded to
the interview questions by e-mail due to exceptional circumstances. Three facilitators (i.e., one per community site) were interviewed by telephone since they were not always present at the community sites. Facilitators were selected based on the information provided by the program managers (i.e., convenience sampling). They answered questions related to the program theory (e.g., goals, objectives, expected magnitudes), process, and implementation (e.g., satisfaction, number of instructors, frequency of the course delivery, data shared with CSC).

Scientific level

Three researchers from the Canadian Driving Research Initiative for Vehicular Safety in the Elderly (Candrive, 2002) took part in this study. Telephone interviews were carried out with these experts. The strategies of judgmental sampling and convenience sampling were applied for the selection of these experts, as they were chosen based on their expertise in mature driving and their geographical location. The questions concerned the theory of the program, such as needs of older drivers in that region and plausibility of the expected outcomes.

Review of Literature

A review of the literature on older drivers was carried out to determine what services and activities were more likely to increase/maintain older drivers’ safe-driving habits. The following electronic databases were searched: PsycINFO, Medline, Google Scholar and Cochrane Database of Systematic Reviews, using the keywords “older drivers,” “driving,” “retraining,” “education,” and “driving behaviours.” Articles retained related to driver retraining programs aimed at improving driving outcomes and were published no earlier than 1990. The nature of the program (e.g., cognitive training, psychoeducation, and physical activity), its delivery, and the expected/observed outcomes were extracted from the articles. Table 1 summarizes the methodology and data sources used for each evaluation question.

Data Analysis

Given the amount of information that the data-collection methods generated, patterns of responses were identified for each evaluation question. Themes were developed for each evaluation question by the three evaluators to ensure that the EA would produce objective analytic conclusions (Wholey et al., 2004). When evaluators disagreed on
The Canadian Journal of Program Evaluation

Table 1
Summary of the Methodology Used for Each Evaluation Question

<table>
<thead>
<tr>
<th>Question</th>
<th>Data Source</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are the objectives and goals of the 55 Alive program clear to stakeholders?</td>
<td>Experts, Facilitators, Program manager of community sites, Program manager CSC, Official documents, Students</td>
<td>Phone interviews, Phone interviews, Face-to-face interviews, Focus groups and telephone interviews</td>
</tr>
<tr>
<td>Are the objectives of the 55 Alive program well defined?</td>
<td>Experts, Program manager of community sites, Program manager CSC</td>
<td>Phone interviews, Face-to-face interviews</td>
</tr>
<tr>
<td>Are the objectives and goals of the 55 Alive program plausible?</td>
<td>Experts, Scientific Literature</td>
<td>Phone interviews, Literature review</td>
</tr>
<tr>
<td>Does the 55 Alive program serve the population for whom it was designed?</td>
<td>Experts, Facilitators, Program managers from the community sites, Students</td>
<td>Phone interviews, Phone interviews, Phone interviews, Demographic questionnaire</td>
</tr>
<tr>
<td>Are the 55 Alive program activities being implemented by the community stakeholders as designed by CSC?</td>
<td>Facilitators, Facilitators' manual developed by CSC</td>
<td>Phone interviews, Observations and document review</td>
</tr>
<tr>
<td>Are the stakeholders satisfied with the 55 Alive program?</td>
<td>Facilitators, Program managers of community sites, Students</td>
<td>Phone interviews, Focus groups and telephone interviews</td>
</tr>
<tr>
<td>Does the 55 Alive program have the capacity to provide data for an evaluation?</td>
<td>Program managers of the community sites, Facilitators</td>
<td>Phone interviews, Phone interviews</td>
</tr>
</tbody>
</table>
the data, to identify whether the evaluators associated them with the same general themes. This statistical measure of inter-judge agreement was considered more useful than percent agreement, because it corrects for chance (Howell, 2002; Stemler, 2001). All of the kappa values were found to be substantial or greater ($\kappa \geq 0.60$). Evaluators followed the benchmarks suggested by Landis and Koch (1977, as cited in Stemler). For the quantitative information generated by the demographic questionnaire, descriptive statistics (e.g., frequencies and means) were calculated using SPSS 15.0.

RESULTS

1. Are the Objectives and Goals of the 55 Alive Program Clear to Stakeholders?

To determine whether the objectives and goals of the program were clear to stakeholders, program managers (i.e., at CSC and community sites), facilitators, students, and Candrive experts were consulted. Specifically, the themes generated from the data collected were compared to the objectives stated in the 55 Alive publicity materials (i.e., recruitment poster and brochure; see Table 2). It was determined that there are some important similarities. Specifically, the themes Correction of bad habits, Increase confidence, and Increase knowledge of road rules were generated from the raw data and are reflected in the objectives in the publicity material (i.e., Identify and correct bad driving habits, Gain more confidence behind the wheel, and Update yourself on traffic laws and new technology). The theme of Sensitive driving environment seems to be captured by the objective of Anticipate the actions of other drivers in the publicity material.

In contrast, some of the themes generated from the data and listed in Table 2 were not reflected in the 55 Alive documentation, including the publicity materials. These themes include: (a) the aging process and how it affects driving (i.e., Increase awareness of age-related decline, Planning for other mobility options, and Increase knowledge of aging and driving), (b) the improvement in driving performance (i.e., Safe driving and Reduce costs associated with motor vehicle crashes), and (c) the licensing process (i.e., Maintenance of driver’s licence and Familiarization with testing). Despite the partial disparity between the objectives stated by the stakeholders and those present in the official program documents, there seemed to be a reasonable degree of agreement.
Table 2
Comparison Between Program Objectives Stated by Stakeholders and Objectives Stated in Publicity Materials

<table>
<thead>
<tr>
<th>Objectives stated by the stakeholders</th>
<th>Objective stated in the publicity materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance of driver's licence</td>
<td>Gain more confidence behind the wheel</td>
</tr>
<tr>
<td>Safe driving</td>
<td>Improve awareness of traffic hazards</td>
</tr>
<tr>
<td>Increase awareness of age-related decline</td>
<td>Update yourself on traffic laws and new technology</td>
</tr>
<tr>
<td>Reduce costs associated with motor vehicle collisions</td>
<td>Anticipate the actions of other drivers</td>
</tr>
<tr>
<td>Planning for other mobility options</td>
<td>Identify and correct bad driving habits</td>
</tr>
<tr>
<td>Increase knowledge of road rules</td>
<td>Voice your concerns in a friendly relaxed environment</td>
</tr>
<tr>
<td>Increase knowledge of aging and driving</td>
<td></td>
</tr>
<tr>
<td>Correction of bad habits</td>
<td></td>
</tr>
<tr>
<td>Increase confidence</td>
<td></td>
</tr>
<tr>
<td>Sensitive driving environment</td>
<td></td>
</tr>
<tr>
<td>Familiarization with testing</td>
<td></td>
</tr>
</tbody>
</table>

2. Are the Objectives of the 55 Alive Program Well Defined?

Generally, the program manager of CSC, program managers of community sites, and Candrive experts had difficulty expressing a magnitude of expected change. Some of our informants, however, estimated the magnitude of change for some objectives (e.g., 30% increase in knowledge, 70% increase in confidence, and 60% increase in knowledge of road rules). Nevertheless, it was impossible to verify these stated magnitudes because benchmark data are unavailable from CSC and the scientific literature. In terms of an expected time frame, most data sources indicated that changes that result from the program would occur immediately after completion, although some stated that it could take years. Again, it is difficult to confirm this time frame as benchmark data are unavailable. It therefore seems that the objectives are not well defined in terms of magnitude and time frame, and that this finding is consistent across stakeholders.

3. Are the Objectives and Goals of the 55 Alive Program Plausible?

To assess the general plausibility, two overarching objectives were explored. Specifically, the Candrive experts were asked about: (a) whether a six-hour knowledge-based class had the potential to change the driving behaviour of older adults and (b) whether such a change would decrease overall crash-rate. The data strongly indicated that experts believed that such a six-hour class would not improve driver
safety. Additionally, two experts indicated that such an outcome would be difficult to assess due to the sheer number of variables involved in a motor-vehicle crash. These responses were consistent with our literature review, which generally found no reduction in at-fault crash-rate among course completers (e.g., Nasvadi & Vavrik, 2007).

4. Does the 55 Alive Program Serve the Population for Whom It Was Designed?

Facilitators and students were asked to identify the characteristics of students that typically take the 55 Alive course. Several themes were identified including the fact that many students were of older age (i.e., over 60 years), many approaching the age of 80 (i.e., mandatory relicensing age in Ontario), English speaking, many had above-average income (between $30,000 and $40,000 annually), and there were no visible minorities. The demographic questionnaire responses mirrored the same themes identified by students and facilitators. These themes indicate that students who take the 55 Alive course are not representative of the general population of older adults over the age of 55 years (Statistics Canada, 2006). These results may indicate that recruitment methods currently target certain populations over others. Specifically, visible minorities, French-speaking individuals, lower- and average-income individuals, and individuals between the ages of 55 and 60 years are underrepresented in the 55 Alive course. Finally, the course participants indicated that they drove approximately 10,000 km last year. Interestingly, the literature indicates that older adults most at-risk of being involved in a motor-vehicle collision are those who drive less than 3,000 km per year (Langford, Methorst, & Hakamies-Blomqvist, 2006). While CSC states that this program is intended to serve all driving Canadians aged 50 and above with the goal of improving road safety, this finding suggests that only a sub-group of Canadians participated. Older drivers most at-risk of being involved in a motor-vehicle collision were not represented in our sample.

5. Are the 55 Alive Program Activities Being Implemented by the Community Stakeholders as Designed by CSC?

This question was evaluated through observations during site visits and interviews with facilitators. Generally, facilitators stated that they followed the instructor’s manual as much as possible. However, they felt it was important to tailor the information to the students’
needs, and stated that, at times, they were forced to omit some material due to time constraints. The observations of the evaluators were scored against the criteria of the implementation grid. This procedure permitted evaluators to determine whether the courses were representative of a 55 Alive course as described by CSC. When more than one rater was present at observation, Kappa coefficients were calculated to make sure that the level of agreement between the two raters was adequate. The resulting Kappa coefficients demonstrated a perfect or substantial level of agreement (ranged from $\kappa = 0.6$ to $\kappa = 1.00$). Both courses offered at Site 1 and Site 2 fulfilled sufficient implementation requirements to be considered a 55 Alive course.

Nevertheless some important discrepancies between the CSC design and the observations are to be noted. First, the facilitators quickly covered the types of driving decisions associated with the aging process, and no homework was given at Site 1 as the course was offered in a full-day session. This was surprising, considering that CSC recommends that the course be given in two half-day sessions. Moreover, at Site 2, students were encouraged to purchase *The Official Ontario Driver’s Handbook* (Ontario Ministry of Transportation, 2007), but not at Site 1. While CSC does not require the use of the handbook, this additional element was found to be particularly useful to students. The facilitator at Site 1, a professional driving instructor, enhanced the course content based on personal experience. Finally, although the CSC advertises that the 55 Alive course is a six-hour refresher course and both facilitators and program managers stated that their programs had a duration of six hours, on observation both courses were found to be significantly (45 minutes) less than this time. This could be explained by the speed of presentation, which a number of participants and the evaluators found to be too fast given the sheer quantity of material to be covered.

6. Are the Stakeholders Satisfied with the 55 Alive Program?

During the phone interviews or focus groups, program managers, facilitators, and students rated their satisfaction with the course on a 5-point scale, where $1 = \text{very disappointed}$ and $5 = \text{completely satisfied}$. Site 1 was consistently rated lower than the two other sites ($M = 3.47$, compared to $M = 4.43$ for site 2 and $M = 4.20$ for site 3). Some students stated that the course material was covered too quickly, the facilitator spoke too softly and some sections were completely missed. Overall, the stakeholders as a group were satisfied with the 55 Alive program ($M = 4.03$) but also found that the program could be greatly
improved. There were several recurring improvement themes, including upgrading material more often, improving audiovisual equipment, increasing speech volume of facilitator and participants, covering more specific situations within the community, covering all course material in greater detail, more time for questions and more participation, and including the Ontario Driver Handbook.

7. Does the 55 Alive Program Have the Capacity to Provide Data for an Outcome Evaluation?

For outcome evaluations, data should be collected relative to the anticipated educational and behavioural outcomes of the program (e.g., improved knowledge, safer driving behaviours). However, our findings revealed that at Site 1, no data were collected from students, except relating to advertising (i.e., where they had heard about the course). At Sites 2 and 3, personal information (i.e., name, address, and licence) was collected. The facilitators at Sites 2 and 3 also use a pre/post measure. This measure is both qualitative and quantitative and asked questions about: (a) personal objectives and goals, (b) whether those were achieved at the end of the course, and (c) how essential a car was in the participant’s everyday needs. Site 2 also included a post-course knowledge test and participants were encouraged to complete the test and review their answers, however no data was collected. At Site 2 all of this information was stored in its original paper format, but at Site 3 the information was stored in a computer system. Additionally, both Sites 2 and 3 stated that they sent participants’ personal information to CSC, but not the pre/post measure data. It was determined that Sites 2 and 3 have the potential to provide data for an outcome evaluation if (a) a standard computerized record of the pre/post tests is put into use and (b) valid measures of the outcomes in the logic model are included. All three program managers stated that they do not share the information collected with other community centres and they do not meet to discuss issues relevant to the 55 Alive course. Similarly, CSC does not collect outcome data.

DISCUSSION

Overall, the stakeholders noted that they were satisfied with the program. The results indicated that the goals and objectives of the program are somewhat clear to the stakeholders. Nevertheless, this EA identified that there was incongruence between the objectives and goals mentioned by the stakeholders of this program and the official
program documents. Clarification of the official program objectives with all of the stakeholders is recommended.

The results from the demographic questionnaires, interviews, focus groups, and site observations indicated inconsistencies between the characteristics of actual attendees and the characteristics of individuals in need of the 55 Alive course. In order to attract a more diverse population, recruitment efforts may be expanded to target a more representative group of potential participants. This could be accomplished by advertising via radio and television. Also, as recommended by one stakeholder, the program could be promoted through a municipal effort, much like first-aid courses. The latter point emphasizes the collaborative aspect of EA, as users of the program are fundamental in generating recommendations for the program.

Currently, there is no benchmark to identify the magnitude and timeframe of the course objectives. This could be directly linked to the lack of data collection by CSC on the intended outcomes. For this reason, CSC might benefit from an outcome monitoring process. The data could be collected relative to the behavioural outcomes linked to these goals and objectives. Since it was revealed that few sites collect data, CSC would need to ensure that sites are aware of the measures to be used. Next, a database must be developed to accumulate the data over successive courses and across community sites. The results of this evaluation also revealed that community sites that collected data did not disseminate their results to other stakeholders. These findings emphasize the need for CSC to foster ongoing communication with community sites. The results also highlight the usefulness of this EA in planning future outcome evaluations as described in Leviton, Kettel Khan, Rog, Dawkins, & Cotton (2010). In other words, the observations of this EA determined the steps that need to be completed if future evaluators wish to conduct relevant and conclusive outcome evaluations of the 55 Alive program.

The findings from the expert interviews suggested that it is unlikely that the program would increase driving safety by actually reducing automobile crashes among older adults. It is recommended that the scope of the program be reduced to focus strictly on increasing driving knowledge. Nevertheless, if CSC aspires to maintain road safety in Canada, it is recommended, based on the scientific literature, that an on-road component be added to the program. Specifically, a trained driving instructor could accompany students in a vehicle and provide feedback based on their performance for one hour, as was done by
Marottoli and colleagues (2007). In fact, the literature suggests the benefit of on-road training in addition to psycho-education in improving the driving performance of older drivers (Bédard et al., 2008). While this complementary component may be associated with several disadvantages (i.e., added financial and human resources demands), the addition of an on-road component may better align the program with the mandate of CSC as well as with evidence-based practices of driver retraining. This recommendation is also congruent with the comments of Fixsen, Naoom, Blase, Friedman, and Wallace (2005) indicating that education alone is unlikely to modify behaviours. Specifically, Fixsen et al. contend that training without a form of coaching or feedback is ineffective. The on-road component would provide such coaching. Additionally, these findings indicated that it was worth using a comprehensive definition of EA, as the EA not only identified shortcomings in the theory of the program as proposed by Wholey (1979), but also generated ideas (e.g., on-road component) to solidify the logic model, which is in line with Smith’s (1989) formative conceptualization of EA.

The results of this EA revealed that the program was implemented by the community stakeholders approximately as designed by CSC. However, some fidelity issues were noticed. The facilitators delivered the program in various ways (i.e., content and dosage). It is plausible that this variability between sites is a variable that could mediate the intended behavioural outcomes of the program. The research of Bédard et al. (2008) supports this proposition in that the behavioural outcomes varied across sites. This finding further suggests that the facilitators play a crucial role in the program’s success and stresses the value of triangulation.

The results of this EA indicate that the program is valued by participants, facilitators, and program managers, as indicated by their high level of satisfaction. In the community sites that were evaluated, the program is generally implemented in accordance with CSC guidelines. However, measurable outcomes were unclear, an ambiguity that seems to originate from an inadequately defined theory of the program. These results could explain the lack of significant improvement in driver safety found in the previous 55 Alive evaluations. The findings also emphasize the importance of having a multi-site approach when conducting evaluations. Evaluating a single site might lead to erroneous conclusions about the program, especially when different individuals are in charge of delivering it. Overall, the results of this evaluation corroborate previous claims that an EA is an activity worth
investing in prior to engaging in a summative evaluation (Rutman, 1980; Wholey, 1979) as it is useful for the organization and, as we discuss in the section below, for evaluation research as a whole.

IMPLICATIONS FOR THE EVALUATION KNOWLEDGE BASE

First, EAs are rarely published in peer-reviewed sources, which limits the available knowledge base pertaining to their usefulness (Smith, 2005, as cited in Thurston & Ramaliu, 2005). This is particularly detrimental to the use of an evaluation because we found that the EA approach was crucial in fostering openness to program evaluation by program managers who are also the program developers. EA relies heavily on program developers, as it requires that the program theory and design be assessed. We found that the formative flavour of this EA facilitated our working relationship with this stakeholder group, allowing us access to original documentation, course manuals, and relevant financial information. They were also enthusiastic about the evaluation results and recommendations, which will most likely augment the utilization of our evaluation. That level of engagement may not have existed had we approached them with an outcome evaluation. Developers of more mature programs (i.e., those that have existed for several years and have been widely implemented), such as the 55-Alive, may feel threatened by the potential negative results of outcome evaluations, which can have a direct impact on their viability. We argue that stakeholders saw the EA as a more collaborative endeavour in comparison to how a summative evaluation may have been perceived; in our experience, EA enhanced the use of this evaluation. Specifically, while the program manager of CSC was unaware of the outcome evaluations conducted previously on the 55 Alive, this stakeholder was actively involved throughout the process and the dissemination of the final report.

Similarly, some authors suggest that EA should be conducted before the program becomes operational (Evaluating Socio Economic Development, 2009). Our findings suggest that EAs are also useful evaluations with mature programs. However, our experience revealed that when an EA is conducted at a later programming stage, evaluators need to assess the openness of stakeholders to fundamental changes that might be costly. Compared to an outcome evaluation, an EA is more focused on theory and process, which means that recommendations might include important changes in the manner in which the program is conceptualized and delivered. As an example, in our evaluation, we recommended that an on-road component be added to
the program if the stakeholders wish to maintain the current scope of the program.

Additionally, the careful selection of our EA framework based on Smith’s (1989) 10-step model (see Trevisan & Huang, 2003) permitted us to evaluate the program as a whole (i.e., from program theory to the anticipated outcomes). This inclusion necessitated the incorporation of a variety of stakeholders. Specifically, we took into consideration those that developed the program (i.e., CSC), those that implemented it (i.e., various community centres), and those that received it (i.e., course participants). This inclusion was key because our most striking findings were those that emerged when the responses of various categories of stakeholders were compared. Thus, EA allowed us to obtain a more holistic snapshot of the program. Had we not examined multiple levels of the program, we would have neglected the fact that the program was not being implemented in the same manner across sites or in the way that it had been designed. Clearly, Smith’s (1989) model (as cited in Trevisan & Huang, 2003) helped generate a more accurate portrait of the health of the program.

As a final point, EAs are fundamental as they inform the evaluator about the level of integration of the knowledge base of the particular issue being assessed (e.g., driver retraining programs and older drivers). Specifically, in this EA we performed a comprehensive literature review on older drivers and driver retraining programs in order to address questions related to the program theory. We realized that the scientific literature on older driver retraining programs was mainly concerned with outcome research (i.e., whether the program improved driving behaviours) rather than the program theory and process. Additionally, the research articles were not situated within an evaluation context as they neglected to refer to fundamental program evaluation principles. Based on our evaluation, it seems that the evaluation literature and studies on the aging driver are not yet integrated. By querying scientists in the field of aging and driving on the basic program theory of the 55 Alive program, we gained valuable insights into the achievability of the outcomes presented in the logic model. Obviously, there is a considerable benefit to involving scientists specializing in a particular area in the program evaluation process; ultimately, their research should serve as the foundation for empirically-based programming. Finally, by including those scientists in evaluation, we may enhance their awareness of evaluation principles and theory thereby improving their outcome research. In other words, with this inclusion, we will come closer to evaluating in the right lane.
ACKNOWLEDGEMENT

The authors would like to thank Dr. Robert Flynn at the Centre for Research on Educational and Community Services of the University of Ottawa for his guidance and thoughtful comments throughout the research process.

REFERENCES


**Mélanie Joanisse** is a Ph.D. candidate in clinical psychology at the University of Ottawa. Her research investigates the impact of psychosocial variables on the driving performance of older adults and is supported by the Social Sciences and Humanities Research Council of Canada (SSHRC).

**Arne Stinchcombe** is a Ph.D. candidate in experimental psychology at the University of Ottawa. His research focuses on the driving challenges of older adults and is supported by the Canadian Institutes for Health Research (CIHR).

**Stephanie Yamin** is a Ph.D candidate in clinical psychology at the University of Ottawa. Her research focuses on the driving performance of older adults with early dementia with Lewy Bodies and early Alzheimer’s disease and is supported by the Canadian Institutes for Health Research (CIHR).