

★ ★ ★ ★ Teen Driver Safety

Review of the Literature on Driver Education Evaluation 2010 Update

Lawrence Lonerio
Northport Associates

Dan Mayhew
Traffic Injury Research Foundation

Teens have the highest crash rate of any group in the United States.



U.S. Department of Transportation
**National Highway Traffic Safety
Administration**



ACKNOWLEDGEMENTS

This report is part of the Large-scale Evaluation of Driver Education Project being conducted by the Traffic Injury Research Foundation and Northport Associates for the AAA Foundation for Traffic Safety. The study is sponsored by the AAA Foundation for Traffic Safety; National Highway Traffic Safety Administration; Centers for Disease Control, National Center for Injury Prevention and Control; and Manitoba Public Insurance.

Founded in 1947, the AAA Foundation in Washington, D.C. is a not-for-profit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, from individual members, from AAA-affiliated insurance companies, as well as from other organizations or sources.

This publication is distributed by the AAA Foundation for Traffic Safety at no charge, as a public service. It may not be resold or used for commercial purposes without the explicit permission of the Foundation. It may, however, be copied in whole or in part and distributed for free via any medium, provided the AAA Foundation is given appropriate credit as the source of the material.

The opinions, findings, conclusions, and recommendations expressed in this publication are those of the authors and are not necessarily those of the AAA Foundation for Traffic Safety, any individuals who peer-reviewed this report, or advisory group members. The AAA Foundation for Traffic Safety assumes no liability for the use or misuse of any information, opinions, findings, conclusions, or recommendations contained in this report.

If trade or manufacturer's names are mentioned, it is only because they are considered essential to the object of this report and their mention should not be construed as an endorsement. The AAA Foundation for Traffic Safety does not endorse products or manufacturers.

1. INTRODUCTION

In North America the term *driver education* has usually been applied to programs for beginning drivers consisting of both “theory” instruction in the classroom and practical training in a vehicle. Traditional programs have taken place in a single stage, before the driver becomes licensed. Indeed, one principal purpose of driver education is to prepare beginners for license testing. In many jurisdictions, but by no means all, most beginner drivers receive some instruction from a paid professional instructor (Maycock and Forsyth 1997).

Driver education was widely available in public secondary schools in North America for many years. After the early 1980s, availability has declined in most jurisdictions. This was especially the case in the United States, where high school driver education had been pervasive through the 1970s. In Canada, driver education has always been more diverse. High school driver education has remained strong in some provinces, never existed in others, and has declined in still others in a manner similar to the United States.

Traditional high school driver education of 30 hours in the classroom and 6 hours in the car is still common in the United States, with some notable exceptions. In some Canadian jurisdictions, the “behind-the-wheel” in-car requirements have been expanded to 8 or 10 hours. In most jurisdictions, classroom methods have typically consisted of teacher-centered lectures, with some discussion supported by film and video. Many commercial driver education providers use similar program structures.

While good instruction can facilitate learning of cognitive and psychomotor skills, better knowledge and skills do not automatically lead to fewer crashes. Substantial evidence suggests that more skillful drivers do not necessarily crash less (Lonero et al, 1995; Williams & O’Neill, 1974; Williams, Preusser, and Ledingham, 2009). Improvements in safety probably require safer driving behavior and habits, not just better skills. That is, to be safer, even better-trained drivers must actually drive differently than they might otherwise have chosen to drive. People in general do not always minimize their risks, and good attitudes are often not expressed in actual good behavior. Lasting behavior leading to lower risk performance in all health and safety fields is much harder to accomplish than is generally understood (Lonero and Clinton 1998; Lonero et al. 1994; Lonero, Clinton, and Sleet, 2006; Sleet and Lonero, 2002).

As an effective behavior change intervention, driver education’s potential seems good, at least on the surface. The typical face-to-face instructional setting provides an opportunity for two-way communication and feedback, as well as rehearsal and practice opportunities. It seems that it ought to provide strong support for forming appropriate habits and skills, as well as reinforcing healthy beliefs, attitudes, and values. Indeed, there was evidence of improvement in knowledge and skills in the DeKalb driver education experimental study (Stock et al. 1983). However, the traditional idea that driver education should lead to long-term safer performance assumes more than improved knowledge and skills, and this connection is yet to be established.

Williams, Preusser and Ledingham (2009) provided a cogent summary of the reasons why traditional driver education may have had less effect than expected:

The courses generally are of short duration, and most time has to be spent teaching basic vehicle handling skills. This leaves less time to try to teach safe driving skills. The audience for driver education may also be relatively unmotivated regarding safety, the primary motivation being to learn enough to get a driver's license. Probably the biggest impediment to driver education effectiveness involves the inherent difficulties in affecting lifestyle and developmental factors: the attitudes, motivations, peer influences, and cognitive and decision-making skills that are so influential in shaping driving styles and crash involvement.

The situation for driver education is really no different than that of short-term school-based courses attempting to influence the use of alcohol, other drugs, or tobacco. These health education programs have also largely failed, for many of the same reasons driver education courses have (p.11)

More education is always a popular prescription for improving safety in any context. However, demonstrated effectiveness in improving safety performance solely through educational measures of any form is relatively rare. In most jurisdictions, a small portion of licensed drivers receive some further instruction, commonly referred to as "advanced" or "defensive" driving, often delivered in an employment setting. While evaluations of advanced programs are beyond the scope of this review, these programs have not usually been found to have a measurable safety impact, although they occasionally have shown some effect in reducing convictions. Violator schools, which are widely used as diversions from court and licensing procedures for drivers receiving traffic tickets, have also been shown to be ineffective in reducing crashes (Gebers, Peck, Janke, & Hagge, 1993; Peck and Gebers 1991). Road safety is not alone in having difficulties demonstrating beneficial bottom-line effects of education. Other health and safety fields share this difficulty (Lonero and Clinton 1998; Lonero et al. 1994).

While driver education traditionally meant instruction only before the new driver was licensed to drive independently, a less-common but potentially important form of instruction is marketed to drivers after they are licensed to drive independently. The distinction between beginner driver education and "advanced" training has become somewhat blurred. A need has long been recognized for additional instruction after a driver has mastered the basics (Lonero et al, 1995) In a few jurisdictions, such as Michigan, Finland, and Luxembourg, new drivers are required to take a second stage of training after they have been driving as licensed drivers for a short period of time (e.g., Glad 1988; Keskinen, Hatakka, and Katila 1998; Shope and Molnar 2003). Some safety benefits of these second-stage programs have been observed, although in a limited range of evaluations.

The aims of this review are to provide a richer understanding of driver education evaluation, as well as perspectives on how beginner driver education evaluation can best be improved in the context of driver education policy, program planning, and

program management. A literature review of evaluation of beginner driver education was included as an appendix to the 2006 AAA Foundation for Traffic Safety document *Evaluation of Driver Education: Comprehensive Guidelines* (Clinton and Lonerio, 2006). The present paper extends and updates the earlier review in light of recently published literature to complement the ongoing *Large Scale Evaluation of Driver Education Project*. The remainder of this paper presents:

- An overview of the background and context for driver education and evaluation;
- A detailed examination of the methods and findings of recent evaluations and some important older evaluations;
- A critical discussion of the methods and conclusions of recent reviews of evaluations; and
- A discussion of the limitations and implications of the evaluation literature for driver education program practice and development.

Young Novice Drivers

When discussing driver education, it is important to keep in mind the common difficulties of young novice drivers that driver education hopes to mitigate. In all motorized jurisdictions around the world, young, inexperienced drivers have much higher crash rates than older, more experienced drivers. Both age and lack of experience contribute to the excess risk (McCartt, Mayhew, Braitman, Ferguson, and Simpson 2008). The crash rate per mile of U.S. 16-year-olds is 10 times the rate of experienced adult drivers (Williams 2003). Crash risk also declines rapidly over the first few months and the first few hundred miles of driving (Mayhew, Simpson, and Pak 2003; McKnight and McKnight 2003). Sixteen-year-olds have almost three times the crash rate of 18-year-olds (Evans 1987). While crash risk drops rapidly, it takes a very long time to level off, requiring as much as 10 years of driving before reaching mature rates (Evans 1991).

Limited skills and abilities contribute to the elevated crash risk of young novice drivers. They are less able than experienced drivers to control attention, scan the environment effectively, detect potential hazards early, and make tough decisions quickly. They perceive less risk in some specific violations and high-risk situations, but they may perceive more risk in certain lower-risk situations. Skill deficits, however, are not necessarily the only problem.

Young novice drivers as a group also tend to raise their risk through seemingly deliberate choices. At least some of them drive too fast and too close to others, accept small gaps in traffic, have unrealistic confidence in their own abilities, and leave inadequate safety margins (Boyce and Geller 2002; Lonerio et al. 1995). Of course, some risky choices that seem deliberate may result from weak ability to anticipate and perceive risks.

Whether the key to young driver risk is weak driving skills or immature decision making has long been debated. Wilde (1994) pointed out that: 1) the excess risk of new drivers is an international phenomenon; 2) the excess risk holds true both per distance driven

and per person; and 3) it is due to both inexperience *and* immaturity. On the other hand, recent research suggests that weak skills due to inexperience might be most important in the early, very high risk driving, as the risk declines substantially over the first few months (Mayhew, Simpson, and Pak 2003). Presumably age-related maturity develops over a longer timeframe and cannot be primarily responsible for the rapid change in crash risk over the first few months of driving, when driving experience builds rapidly.

James and Scott McKnight (2003) studied the records of non-fatal crashes of young novice drivers. They concluded that collision reports typically evidence simple mistakes, seemingly consistent with inexperienced skill failures rather than extravagant risk taking. The most prominent errors included lack of visual search prior to turning left, not watching the car ahead, driving too fast for conditions, and failing to adjust adequately for wet road surfaces. These error patterns did not change across the 16-19 age range.

If young drivers' non-fatal crashes are precipitated by relatively minor errors, one might reasonably expect to see a different pattern of errors in fatal crashes, which differ in many ways from the patterns of less severe crashes. Many young driver fatal crashes involve a single vehicle. A study of U.S. Fatality Analysis Reporting System (FARS) data for the State of Colorado (Gonzales et al. 2005) suggests a much higher incidence of violations in young driver fatal crashes than in mature driver fatal crashes (e.g., speeding – 1.9 times higher, driving recklessly – 4.8 times higher). The researchers also found lower incidence of some risk factors, such as alcohol impairment and adverse weather conditions. This study supports the expectation that young driver fatal crashes are different from older drivers' fatal crashes and from young drivers' non-fatal crashes.

¹ Waiting until another study with larger numbers of cases is conducted will be necessary to validate these differences through fatality data. It may well be that fatal crashes and common minor crashes are typically different in etiology and require different theoretical approaches to prevention through education and other countermeasures.

Inadvertent errors and unsafe choices probably both contribute to young novice drivers' excess risk, albeit perhaps not in the same proportions for differing severities of crashes and at different times in the early driving career. This implies that both error avoidance and safer choices should be effectively addressed in driver education and serve as evaluation targets for longer-term driver education evaluation.

2. PROGRAM CONTEXT FOR DRIVER EDUCATION EVALUATION

Driver education has long had been mandated to address all possible aspects of the tragically high crash risk of young novice drivers. Courses for beginners have long been a popular and convenient means of achieving independent mobility, important for both

¹ The differences that were found, however, although seemingly substantial, do not appear to be statistically significant. Although the report does not address significance of the odds ratios found, it does provide 95% confidence intervals, which seem to bracket all the odds ratios presented.

young people and their parents. Driver education has strong “face validity” as a safety measure. Parents think it makes their children safer drivers (Fuller and Bonney 2003, 2004; Plato and Rasp 1983).

A concise and detailed history of driver education can be found in a paper by Nichols (2003) prepared for the 2003 U.S. National Transportation Safety Board (NTSB) hearings on driver education, available in the NTSB (2005) report on driver education. The NTSB hearings and recommendations were perhaps a turning point toward more organized, systematic, and evidence-based development, after a long period of relatively little centralized support or direction for driver education R&D. A recent history is also provided by Williams, Preusser, and Ledingham (2009).

In recent years, there have been changes in how driver education is delivered. There are rapidly developing trends toward changes in instructional method and program delivery. Traditionally, all driver education activities involved face-to-face interaction between instructor and learner, although classroom instruction was often supported with film and video media, and sometimes simulators. More recently self-instruction, computer-based instruction, simulation, and even web-based instruction have become prevalent, particularly in parts of the United States.

Most of the changes have been the result of technology “push” rather than having been pulled by pedagogical, epidemiological, or evaluation research on driver education and its target audiences. Nevertheless, as a result of recent changes, driver education is now quite diverse in its delivery and organization. As well as traditional high school driver education, evaluators must recognize the growing importance of driving schools, home schooling, and computer- and web-based instruction.

Even though content has changed less than delivery, there is also greater diversity in content now. This diversity complicates answering general evaluation questions such as, “Does driver education work? How can its effects be improved?” This review examines how these questions have been asked and answered in the past to provide a more in-depth understanding of driver education evaluation and to identify evaluation implications for improving driver education policies and practices in the more complex and diverse future.

The key development in the wider field of driver education in North America in recent years is the NHTSA-sponsored *Novice Teen Driver Education and Training Administrative Standards* (National Driver Education Standards Project 2009). A consultant-led process involving a large group of expert advisors produced a new standard or guideline for the states’ driver education programs. The purpose of the voluntary standard is to advance the best practices in driver education and to improve weaker programs in a number of states. It addresses administrative matters, suggesting that there should be a lead agency in charge at the state level – not always the case in the US now – with power to approve programs according to the state’s own standards. Coordination of driver education with driver licensing is recommended as is better integration of driver education with graduated driver licensing (GDL).

Principal content emphasis is on longer courses (45 hours in-class/ 10 on-road) and use of a nationally recognized model, such as ADTSEA, NIDB, or DSAA for each state's program. Simulation and other computer-based methods are recognized as legitimate teaching tools, but they should not to be substituted for basic in-car instruction.

Evaluation is addressed briefly in the Standard. Under Management, Leadership, and Administration, each state should:

- Have standardized monitoring, evaluation/auditing, and oversight procedures to ensure that every driver education and training program uses a curriculum with written goals and objectives (1.1.4.)
- Require all public and private driver education and training providers to report program data to the designated state agency so that periodic evaluations of the state's driver education and training programs can be completed and made available to the public. (1.1.16)

Under Education/Training, each state should:

- Require a course provider to conduct valid post-course evaluations of driver education and training programs to be completed by the students and/or parent for the purpose of improving the effectiveness of the program (a resource for help in conducting these evaluations is the AAA Foundation for Traffic Safety). (2.1.5)

The last of these refers to the 2006 AAA Foundation for Traffic Safety document *Evaluation of Driver Education: Comprehensive Guidelines*.

While the new NHTSA administrative standards are relatively weak on evaluation research, NHTSA's *Highway Safety Program Guideline No. 4: Driver Education* prescribes a rather comprehensive approach to evaluation and program development. The Guideline indicates, "Evaluation should be used to revise existing programs, develop new programs, and determine progress and success" (p.1).

Under Program Evaluation and Data, the Guideline reads:

The SHSO (State Highway Safety Office), in collaboration and cooperation with the State agencies responsible for driver education and training, should develop a comprehensive evaluation program to measure progress toward established project goals and objectives and optimize the allocation of limited resources. The State should promote effective evaluation by:

- Supporting the analysis of police accident reports;
- Encouraging, supporting, and training localities in process, impact, and outcome evaluation of local programs;
- Evaluating the use of program resources and the effectiveness of existing countermeasures for the general public and high-risk populations; and

- Ensuring that evaluation results are used to identify problems, plan new programs, and improve existing programs. (p.6)

It is clear that the context for driver education and for evaluation in the field is rapidly changing, after long periods of relative stasis.

3. DRIVER EDUCATION EVALUATION OVERVIEW

In most types of education, a course is considered successful if learning objectives are met at the end. How graduates choose to use their learning is considered beyond the responsibility of the course. Driver education's assignment has been much tougher, because it has been expected to produce improved subsequent driving behavior *and* measurable effects on crashes. Crashes have complex causes, some aspects of which are outside the driver's control; and even among high-risk drivers, crashes are rare events, so it is difficult to measure small effects of driver education programs.

Compared to the public health and education fields, beginner driver education has seen relatively few evaluations, and these have typically assessed short-term, direct-safety impacts. These impacts have usually been defined as total subsequent reported crashes of graduates compared to new drivers who learned to drive in other ways, and who passed the same licensing tests. Numerous evaluations have compared different forms of formal driver education, with some including a "no-treatment" control group, consisting of new drivers who learned to drive through their family without any professional instruction.

Research design of driver education evaluation has generally involved one of three basic approaches. These are:

- **Experimental studies** that randomly assigned drivers to various training conditions and compared subsequent crash rates and other measures;
- **Quasi-experimental studies** that observed differences between self-selected driver education students and those who learned to drive in other ways; and
- **Ecological studies** that have considered impacts on crashes following jurisdictional changes in requirements or support for driver education.

The largest experimental study and most influential driver education evaluation ("DeKalb") involved randomly assigning 16,000 U.S. high school student volunteers to three groups – intensive training, minimal training, or no formal driver education. The results failed to show a dramatic, long-term benefit of a special course (Stock et al. 1983), and reactions to the results had profound effects on driver education. In the United States, driver education market penetration peaked in the early 1980s with about 80% of new drivers being formally trained. Afterward, however, many high school driver education programs were dropped. New Jersey schools offering driver education, for example, dropped from 96% to 40% between 1976 and 1986 (Simpson 1996). Although it certainly has been blamed, it has not been clearly demonstrated whether the DeKalb

results in effect caused the decline in U.S. high school driver education, or whether it served as a convenient excuse for widespread budget cutting at the time.

Other studies than DeKalb have also found that driver education failed to produce a measurable change in crash rates. Although positive findings have also been found, many members of the safety research community have come to believe that “driver education does not work.” This conclusion raises questions as to how such a counter-intuitive situation might be possible. However, given the limited scope of beginner training and its position at the very start of a long learning curve, driver education effects may be overshadowed by other experiences, overconfidence, increased exposure to risk, and relaxed parental supervision. Since so much of a new drivers’ learning takes place after licensing, potentially beneficial effects of traditional driver education may be offset by other influences. And as researchers have also suggested, driver education in the past may not have provided the best possible content in the best ways (Mayhew and Simpson 1997).

Unfortunately, as will be discussed in more detail later, evaluations in driver education have also been rather unsystematic and limited in quantity and quality. Even the randomized controlled trial (RCT) experiments suffer methodological problems that make their results less than definitive. Some studies used small samples and lacked statistical power to detect modest effects (Engström et al. 2003). A recent Australian quasi-experimental evaluation observed substantial crash differences between training conditions, but, because the numbers of drivers were so small, it could not conclude that the differences were the result of anything other than chance (Haworth, Kowadlo, and Tingvall 2000). Very large numbers of cases are needed, even to assess effects on total crashes, let alone injury or fatal crashes.

Other study design problems also have reduced evaluation benefits for driver education. Most evaluations have failed to look at intermediate student outcomes – that is, the knowledge, skills, attitudes, intentions, or values that had (or had not) been affected by driver education. Ways to improve driver education programs have, therefore, been unclear (Lonerio et al. 1994; 1995). Intermediate outcome measures and survey tracking of behavior during the follow-up period can provide something akin to an “audit trail” of program outcomes, such as knowledge, skill, attitudes, and exposure to risk, which in turn, lead to safety impacts.

Key areas where driver education evaluation has been found lacking include:

- Program theory: Theory in the sense used here means the logic model that justifies thinking a program should meet its goals – that is, why we think it should work. There has been little evaluation of the theory underlying various driver education programs.
- Formative evaluation: This is applying evaluation tools to improve the content and delivery of a program. Traditionally, there has been little formative evaluation of intermediate effects, so it is not clear how well driver education students achieve,

retain, and use desired skills and knowledge. Driver education courses vary greatly in quality, and limited evaluation of program differences has existed.

- Methodological soundness: In the pool of existing evaluations, problems of scope, design, and sampling limit unequivocal conclusions.

Early studies made no effort to control for the ways in which driver education graduates were different from comparison groups other than the type of training each group had received. As a result, these uncontrolled quasi-experiments were not considered credible. To remedy this problem, some later evaluations were designed as RCTs; that is, they controlled for extraneous differences between driver education and comparison groups by random assignment to a training group or a control group (comprised of informally trained novice drivers or those who took a different course).

Overall, scientific evaluation of driver education has been quite limited in quantity, quality, and scope. Beginner driver education is particularly hard to evaluate, at least in terms of safety impacts, because suitable comparison groups are hard to establish. Many earlier evaluations compared groups of young drivers who not only received different forms of driver education, but also differed in other ways that might have affected their driving records and other outcomes. These “extraneous” or “confounding” factors could include location of residence, income, or other important socioeconomic factors. Even when they can be established, comparable treatment and control groups are hard to maintain over time, since assigned or selected groups can have different dropout rates.

4. REVIEWS OF DRIVER EDUCATION EVALUATIONS

The present section briefly discusses the conclusions and limitations of the most significant recent reviews of driver education evaluations.

Past reviews of driver education evaluations have mainly attempted to determine the extent that evaluations indicate whether driver education works. This review has a somewhat broader focus. It provides a perspective on driver education evaluation in the context of program planning and management, and on *how evaluation might be improved* to play a stronger and more positive role in future driver education development. Evaluation is a potentially valuable and crucial tool for improving driver education, but it has not yet achieved its full potential.

Since the mid-1990s, a number of broad reviews of evaluation studies of beginner driver education have appeared (Christie 2001; Engström et al. 2003; Lonero et al. 1994, 1995; Mayhew and Simpson 1997, 2002; Nichols 2003; Siegrist 2003; Smiley, Lonero, and Chipman 2004; Williams et al. 2009; Woolley 2000). Some of the recent reviews have approached driver education evaluation from a health perspective, essentially trying to determine its effectiveness as an injury prevention “treatment.” Others have addressed driver education in conjunction with other forms of driver instruction or as an

adjunct to graduated driver licensing (GDL). A single quantitative meta-analysis of the evaluation literature has been provided by Elvik and Vaa (2004),

Early driver education evaluations were reviewed by Nichols and summarized as follows:

Although there were a few early studies which reported negative results, the majority of studies conducted at the time suggested that: (1) while the effects may be short-lived, driver education students had fewer accidents and violations than non-students; (2) complete courses involving both classroom and behind-the-wheel training were more effective than classroom-only courses; and (3) [High School Driver Education] was more effective in reducing accidents and violations than either parent training or commercial driver training. (Nichols 2003, 20)

Systematic Reviews

Two recent “systematic” reviews of selected, small numbers of evaluations have appeared. Vernick, Li, Ogaitis, MacKenzie, Baker, and Gielen (1999) reviewed nine evaluations that met the authors’ methodological criteria (of 27 evaluations found). The review’s stated intent was broader than most, aimed at finding: 1) whether driver education graduates were less likely to crash or more likely to become licensed to drive; and 2) whether driver education had broader public health effects in lowering community rates of crashes. All but one of the nine studies addressed U.S. high school programs. Five of the studies reviewed were structured as large-scale ecological record modeling studies, including two each from Levy (1988; 1990) and Robertson (1978; 1980). The other four studies were experimental RCTs. Three of the RCTs consisted of one of the original DeKalb experiment reports and two re-analyses of this data. The fourth was an RCT with fewer than 800 subjects assigned to four different training conditions, which, not surprisingly given the small sample, found no significant differences in crashes (Strang et al. 1982). The reviewers concluded that no study that met their design criteria showed a “significant individual or community-level beneficial effect of driver education for high school-aged students” (p. 44). No explanation is offered for disregarding the findings of a significant beneficial effect on fatal crashes by Levy (1990).

Using an even narrower selection basis than the Vernick et al. review, Roberts and Kwan (2004) reviewed three RCT experimental evaluations, all from the early 1980s. They also concluded that no evidence showed safety impacts of driver education. The utility of this review of old and well-known studies is difficult to discern. Neither of the systematic reviews included the RCT study by Dreyer and Janke (1979), which found a positive effect on drivers’ crash records. The narrow orientation of this approach to systematic review seems to limit its applicability to driver education evaluation, since most evaluations would not meet the criteria for inclusion, and RCTs present a special difficulty in evaluating beginner driver education, as will be discussed more fully at the end of this review.

Review of Driver Education in Graduated Licensing

Graduated licensing that delays independent driving has been the principal initiative to address young driver crashes in recent years and has been shown to be effective in reducing crashes. At the Traffic Injury Research Foundation of Canada, Mayhew and Simpson (1997) performed a detailed review of the DeKalb experiment and eight later evaluations of beginner driver education, in the wider context of graduated licensing and other forms of driving instruction. These researchers indicated some positive findings for driver education effectiveness for novice car drivers, including:

- Per-licensed driver analyses of the short-term DeKalb data (Stock et al. 1983);
- Minimal training condition in a long-term follow up in the DeKalb experiment (Smith and Blatt 1987);
- Quebec mandatory training system (Potvin, Champagne, and Laberge-Nadeau 1988);
- Norwegian second-stage training in night driving (Glad 1988);
- Denmark mandatory driver education (Carstensen 1994); and
- Finland mandatory slippery road training for older novices (Katila et al. 1995).

They also pointed out that two U.S. studies in the 1980s showed better safety effects from school-based driver education than commercial schools, but these were uncontrolled quasi-experiments, similar to the early evaluations and have low credibility. However, the econometric modeling study of driver education over 47 U.S. states (Levy 1990), which also showed a positive effect for driver education, was not included in this review. Also not included was the 1979 California RCT study by Dreyer and Janke, which found a difference in effect between two approaches.

Over a wide range of driving instruction, the authors concluded that 7 studies showed a positive effect, 16 showed no effect, and about 7 others showed a negative safety effect. They summarized their findings as follows:

On balance, the weight of the available evidence does not favor the hypothesis that formal instruction has safety benefits. Indeed there is precious little in the way of reliable evidence to show that formal instruction provides safety benefits. As counterintuitive as this may seem, it is difficult to reach a different conclusion in the face of the total body of evidence. (Mayhew and Simpson 1997, 45)

As well as recognizing some positive findings, Simpson and Mayhew's review differed from most by making suggestions for improving driver education. They recommended three key directions for future improvement of driver education:

1. Driver education should be multi-phased;
2. GDL systems should not give a "time discount" for driver education; and
3. Specific changes should be implemented in the content and delivery of driver education.

Mayhew and Simpson (2002) revisited the evaluation literature in a later brief review, which also discussed other recent reviews. Similar conclusions resulted.

Australian and New Zealand Reviews of Driver Instruction

Christie (2001) in Australia recently published an influential and detailed review of evaluations of various forms of driver instruction, including beginner driver education. He reviewed the same studies as Mayhew and Simpson, as well as later ones, including one from Pennsylvania that will be discussed in the section on individual evaluations. Christie provided an analysis of a Tasmanian retrospective records-based study that was unavailable to the present review (Langford 1997, cited in Christie 2001). The Langford study apparently showed positive initial effects of a limited high school course, but later re-analysis of the same cohorts with fuller reporting of records indicated no differences.

Apparently less impressed than Mayhew and Simpson with the limited positive impacts of driver education found in the literature, Christie concluded that no evidence shows beneficial effects of beginner driver education. He reiterated the view that driver education is harmful because it induces earlier licensing. Christie summarized his view as follows:

New approaches to driver training may eventually prove to be useful in reducing casualty accident risk/involvement, but much research and development work remains to be done before one could say that driver training is an effective crash countermeasure. In the interim, other approaches such as increased supervision and graduated licensing for novice drivers and traffic law enforcement for all drivers are likely to make greater and more lasting contributions to road safety. (Christie 2001, 43)

In a later review of road safety education, Christie (2002) also addressed public information and advertising programs for road safety. He concluded they too are ineffective except as adjuncts to legislative and enforcement programs. Again, he suggests that other kinds of programs, including graduated licensing, enforcement, crashworthy vehicles, and “black spot” roadway hazard correction should be given priority. This broader view of road safety management as a context for driver education in all forms is well considered and consistent with earlier recommendations by Lonero et al. (1995).

Christie’s analyses, however, seem to stop short of applying the same standard of effectiveness countermeasures other than driver education. Using standards of proof applied to driver education, most other behavioral safety measures also cannot be shown to be effective for preventing crashes. As the present authors have concluded from reviewing the full range of safety programs aimed at road user behavior, very few behavioral interventions work well in isolation (Lonero et al. 1994). Of course, similar conclusions regarding the need for coordinated programs of behavioral influence have also been drawn in other fields that attempt to change behavior, such as health promotion. (Green and Kreuter 1991; Lonero and Clinton 1998).

Planned and coordinated combinations of influences seem to work, but single-technique approaches typically do not. Unplanned and uncoordinated combinations of influences may even add up to changes in culture and behavior over time. Examples might include the changes in seat belt use and impaired driving, which have occurred despite individual educational and enforcement interventions often showing no effect or short-lived effects. Little or no research has yet addressed these broader areas of safety behavior change, which Christie correctly attempts to bring to the driver education discussion.

Another Australian review (Woolley 2000) concluded that non-skills factors are the keys to resolving road safety problems and that no conclusive link exists between skills-based training and crash involvement. Rather, motivation and risk-taking propensity are more important than any type of skills-based training, and driver education should be developed to address these critical factors.

Difficult choices are presented when the benefit of a critical program, such as driver education, does not appear as expected. Should resources be shifted to other kinds of programs, as suggested by Christie? Or should additional resources be employed to improve the disappointing program to make it effective in the ways expected?

Swedish Review of High School Driver Education

A brief review appeared as a chapter by Inger Engström and colleagues in a wider Swedish review of licensing and other behavioral measures for novice drivers (Engström, Gregersen, Hernetkoski, Keskinen, and Nyberg 2003). This review covered some of the same evaluations as DeKalb and its reanalyses, as well as Australian and New Zealand studies. It also covered some U.S. state program studies that are more like qualitative program reviews than the more typical quantitative evaluation studies. In addition, these researchers reviewed a small evaluation (unavailable to the present authors) of a Swedish 30-hour high school pre-driver education course, and they reported that survey results showed improved knowledge of traffic risks and laws.

The Engstrom review briefly touched on evaluation methodology issues, pointing out the controversy over appropriate measures for assessing driver education effects. Engström did, however, maintain the distinction between safety measures, as typically defined, using only intermediate criteria, such as behavior or attitude measures. Alternative ways of looking for safety effects seem to exist, beyond those that have been used so far in driver education evaluation. Engström et al. also pointed out the common problem of unacceptably small sample sizes, which has plagued the driver education evaluation field from its inception and continues in recent studies.

The authors concluded:

The current approach, as far as evaluated, has not shown the benefits that could be expected from education and training in high schools. It is thus of the highest

importance to develop new strategies since the high school environment provides good possibilities to reach youngsters with safety issues that are not normally covered by driving schools or parents. (Engström et al. 2003, 93)

Engström et al. did not provide an explanation of why they think these possibilities are more likely to succeed than driver education itself, but reiterated the prescription of Woolley (2000) that driver education should have a supportive role in a holistic approach to road safety in schools.

Norwegian Meta-Analysis of Evaluation Studies

Elvik and Vaa (2004) reported the results of an extensive meta-analysis of driver education evaluations from around the world. This approach, which combined the data findings of 16 individual studies, overcomes some of the difficulties with the individual evaluations, such as the common problem of too-small sample sizes. The meta-analytic approach recognizes that no single evaluation is likely to be definitive, and this astute study can be seen as the most definitive summary of the past evaluations.

Unfortunately, the only report on the meta-analysis is a very brief one in the 2004 edition of the Norwegian *Handbook of Road Safety Measures*, which is available in English.

The combined data of all 16 studies indicated that driver education graduates have 1.9% fewer crashes per driver (confidence interval, -3.8%; 0%). The overall difference appears to have been nearly statistically significant, as the combined results had the power to detect a significant difference ($\alpha=.05$) if the result had been 2% rather than 1.9%. Per kilometer driven, there was a 4% lower crash rate for graduates (-6%; -2%), but statistical significance for this difference was not reported.

When the combined results were limited to the experimental studies, however, a different picture emerged. No difference per driver surfaced (+/- 4%). Per kilometer driven, driver education graduates had 11% more crashes (+8%; +15%), again the significance of which was not reported. The authors conclude that the combined evaluation results do not indicate that driver education reduces crashes over the first couple of years of driving.

Elvik and Vaa also examined, briefly, four possible explanations for the generally disappointing findings among driver education evaluations. The first explanation is that the evaluation research is too poor to detect the real effects of driver education. This is refuted by indicating that the research overall is actually somewhat better than the evaluation research typical in most road safety programs. They also suggest that only the poorest studies have found any positive effect, although significant positive effects have appeared even among the favored experimental studies (e.g., Dreyer and Janke 1979).

The second possible explanation for lack of positive findings is that programs evaluated are not good enough. The authors see this as unlikely, however, because the best programs are probably those that have been evaluated. This is plausible, except that

most evaluations took place in the distant past, and most current programs have not been evaluated. The third explanation is that crashes are too insensitive a measure to detect training effects. The authors are able to refute this possible explanation of no effects, as they indicate that the combination of results across of all the 16 evaluations raises the statistical power enough to detect even a 2% crash difference.

The final potential explanation, favored by the authors, is behavioral adaptation –less-skilled drivers taking more care and better skilled drivers taking less. While the meta-analysis results do not directly address this explanation, the authors cite the negative effects found in evaluations of skid training, as well as Gregersen’s (1996) study, which showed that skid training could raise confidence without actually increasing skill. Many researchers have been skeptical of behavioral adaptation, and even if this explanation is accepted, it might raise the question of whether the best current or future driver education can be “good enough” to help overcome such motivational difficulties and, given good enough evaluation, clearly demonstrate an effect on crash rates.

U.S. Feasibility Analysis and Review

Williams, Preusser and Ledingham (2009) published a long-awaited research paper on the feasibility of evaluation of the ADTSEA (American Driver and Traffic Safety Education Association) driver education curriculum. The objectives of this project were to examine previous evaluations, focusing on research design issues, and to determine the feasibility of a comprehensive evaluation of the new curriculum developed by ADTSEA.

The authors’ perspective on design alternatives and guidelines for summative evaluation of safety impacts of the new ADTSEA program are presented. They concluded that the ADTSEA curriculum could potentially be evaluated using experimental designs, quasi-experimental designs, or more naturalistic, ecological approaches. However, it has not yet been implemented “in a way that can be scientifically evaluated, and it would need to be introduced in schools or communities in ways that facilitate its evaluation” (p19).

The ADTSEA program cannot be evaluated presently using any of these possible designs. This is because the program has been implemented in schools or communities in only limited ways...to the extent it is used, it has generally been added to existing driver education programs, often in piecemeal fashion. Individual schools (with their own guidelines) use what they want but rarely the whole program. Given this situation, a fair test of the ADTSEA program is not presently possible (p.17).

The report provides an overview of the history of driver education in the United States. The report also includes a selective review of evaluation studies and earlier reviews, similarly concluding that evaluation research indicates no safety benefit of driver education and even some negative effects due to earlier licensing.

The paper adds a novel perspective on the earlier licensing effects seen in some driver education evaluations, seeing it as a form of self-selection bias.

This debate is highlighted because it is prominent in discussions about the most appropriate way to evaluate driver education programs, specifically whether the analysis should be based on all those assigned to the program (which captures the early licensing effect, plus any effects of driver education) or only on those who become licensed (which supposedly captures the effects of driver education alone). In the DeKalb study, where this issue was critical, students self-selected into the licensed driver subgroup in a way that invalidated analyses based on licensure as reflecting the effects of driver education (p.12).

Williams and colleagues also contributed a strong discussion of the relative merits and practicalities of randomized experimental study designs in evaluation of beginner driver education, pointing out the great difficulties with the method in the DeKalb study. There is also a sobering discussion of the very large sample size requirements for crash-based experimental studies, either randomized or quasi-experimental, to reach statistical significance with moderate effect sizes.

No recognition of any other role or approach to evaluation research is presented in the paper. This paper indicates no consideration of formative aspects of evaluation research or other ways in which evaluation might be used to support development of more effective programs. Rather than discussing how evaluation research might better be structured to support the implementation and development of the target curriculum, it addresses ways the program might be better implemented to suit evaluation needs. These aspects of the paper reflect the tradition of the experimental-psychology oriented research community in driver research, rather than evaluation research as a discipline.

5. INDIVIDUAL EVALUATIONS OF DRIVER EDUCATION PROGRAMS

The more recent individual quantitative evaluations, along with selected older evaluations, are discussed in this section.

The great majority of driver education programs have never been formally evaluated, and most existing evaluations are severely limited in scope, power, and scientific rigor. This section describes selected individual evaluations, which represent either fairly recent work in the field, or older studies of special historical importance. As indicated earlier, the three basic types of studies are:

- **Experimental studies** – students are assigned to different training conditions;
- **Quasi-experimental studies** – naturally occurring groups are compared; and
- **Ecological studies** – assessment of changes in driver training requirements or program differences across different jurisdictions.

These categories are used to group the studies to be reviewed.

Experimental Studies of the Impact of Driver Education

DeKalb County, Georgia

The U.S. DeKalb County Driver Education Project was the most comprehensive experiment in beginner driver education, based on the typical delivery of U.S. driver education in public secondary schools. The DeKalb Project is best known for its impressive efforts to provide improved training and well-controlled experimental evaluation of subsequent crashes over six months, one year, and six years (Lund, Williams, and Zador 1986; Ray et al. 1980; Smith 1983; Stock et al. 1983).

Volunteer high school students were assigned to one of three groups that received different driver instruction conditions. The random group assignment was intended to eliminate self-selection bias, which had troubled earlier attempts to evaluate driver education (Vernick et al. 1999). That is, in the normal course of events, beginner drivers who take driver education courses are different from those who do not, in other ways as well, and these other differences bias any attempt to compare their subsequent driving records. Random assignment to treatments in theory eliminates this bias, even if it introduces some other problems in practice.

A special educational program called the Safe Performance Curriculum (SPC) was developed for the DeKalb study. The new curriculum was based on a comprehensive driving task analysis, and was intended to represent the 1970s state-of-the-art driver education, both in terms of content and methods. The SPC, which was pilot tested and improved before implementation, was a much longer program and more carefully developed than typical driver education curricula of its time. It was strongly oriented toward improving drivers' hazard perception skills, and it de-emphasized the motivational "scare tactics" that had been prevalent earlier.

The SPC consisted of 32 hours of classroom instruction, 16 hours of simulation instruction, 16 hours of driving range instruction, 3 hours of collision evasion instruction, and 3.3 hours of on-road, behind-the-wheel instruction, including 20 minutes at night (Lund et al. 1986).

The effect of SPC training on crashes was compared to two other training treatment groups. A minimal training course called PDL (Pre-Driver License) was intended to represent the minimal training required to pass the licensing test. It was essentially a standard driver education course of 30 classroom hours and 6 driving hours. The other comparison group was a "no-treatment" control group. The young people assigned to this condition were denied the opportunity to receive driver education in the public schools, leaving them to find their own instruction from family members or private instructors. The training the members of the control group actually received was apparently not known.

To overcome an initial resistance among those assigned to the control group and induce them to participate in the experiment, insurance discounts were offered

equivalent to those available to the training group participants. These discounts were available to control group members if they passed a special road test that was used for the other training groups (Stock et al. 1983). The number of the supposedly untrained group who prepared themselves for this rigorous test is unknown, but such thorough preparation would presumably reduce the contrast between the two training groups and the supposed no-treatment control group. This is a clear illustration of the difficulty of applying the randomized controlled trial method to beginner driver education. Unlike the treatment assignment in an ideal RCT, subjects in the no-treatment group were not prevented from receiving some form of driver education. They were able and even encouraged by insurance incentives to respond to their group assignment in ways that could defeat its purpose, and the extent to which they did this was not adequately tracked.

Intermediate learning outcomes were assessed to some degree in the DeKalb study. Testing was carried out, and records were followed for the young volunteers in each of the three groups. SPC-trained drivers showed better on-road skills than the other groups.

SPC graduates that received drivers' licenses also experienced fewer collisions per licensed driver over their first six months of licensed driving. When the data were reanalyzed, however, to examine net benefits for the entire group assigned to the SPC (whether or not they actually took the course), no benefit in collisions was found. The reduction of collisions per licensed driver seen in SPC drivers was offset by earlier licensing, and therefore more exposure to risk, for the young drivers assigned to SPC compared to the groups assigned to the minimal formal training and to no formal training.

After six months, even collisions per licensed driver were no longer better for the SPC group. In a long-term follow-up study of the records of the DeKalb students over six years, both the SPC and minimal curriculum males were found to have significantly fewer convictions, and both males and females in the minimal curriculum group had fewer crashes (6%) than the untrained controls (Smith and Blatt 1987; Weaver 1987). Over the long term, these studies indicated that minimal training led to slightly lower crash rates than did the more extensive and intensive SPC training. This finding was somewhat puzzling, as it seemed to show a delayed effect of training (Ray Peck, personal communication). A later reanalysis of the DeKalb data by Davis (1990) found no differences in the crash rates of the three training groups after the first year following training. Davis also strongly questioned the technical adequacy of the statistical methods used by some of the earlier analyses of the DeKalb data.

Mayhew and Simpson (1997) conducted a detailed review of the original DeKalb study findings and the subsequent analyses by other researchers, and concluded:

Thus despite significant effort, the DeKalb evaluation produced findings that failed to provide evidence of the beneficial effects of formal instruction. Not surprisingly, the equivocal nature of the results has led to substantial controversy

that has had a profound impact on driver education/training. (Mayhew and Simpson 1997, 20).

Disappointment with the findings for the SPC graduates led to withdrawing support and subsequent stasis and decline of driver education in North America for many years (Nichols 2003). Smith (1983) viewed the issue more as one of specific training effectiveness and less as one of engineering safety on a broad societal scale. He contended that collision measures are not the appropriate criteria to assess a program that has a main objective to ensure proper and safe driving performance because collisions are not common occurrences, are valid measures of driver performance only in conjunction with measures of exposure, and do not reflect the full range of driving ability. He recommended adopting an intermediate criterion developed for the DeKalb project's improved curriculum. This measure was based on observed behavior in selected traffic situations. According to Smith, it measures:

... both cognitive and non-cognitive behaviors, observes actual behavior patterns in relation to real-life traffic, and records interrelationships of driver behaviors to changing traffic conditions. ... Such a criterion yields immediate results, is accumulated in a short period of time, identifies proficiencies/ deficiencies in response to real world situations without waiting for people to injure or kill themselves. (Smith 1983, 26)

The DeKalb experiment has numerous implications for the field. It suggests that even carefully planned RCTs can have difficulty in achieving and maintaining assigned treatment groups, particularly in a no-treatment control group. It raises endless possibilities for debate over proper comparisons. If measurable effects of a program exist, how long should they be expected to last before we consider them real or important? Should assignment to the target treatment group be considered the treatment, so that whole assigned groups are compared? If so, aren't the differential effects of the compared programs being diluted by those who are assigned to, but do not actually attend or complete the program?

From a broader program management perspective, the DeKalb experience also points out the limitations of the program development theories prevalent at the time. Progress in programs is rarely made by monolithic developments. More modern theory suggests incremental, continuous improvement approaches would more likely lead to success. Finally, the policy response to the disappointing results was probably not the right one – in retrospect, reducing research and program development resources did not kill driver education but just delayed its development.

Australia and New Zealand

Wynne-Jones and Hurst (1984) conducted an experimental trial evaluation of the Automobile Association (AA) high school driver education program in New Zealand, using random assignment to conditions to eliminate self-selection bias. One group (561 students) was offered a course consisting of 15 hours in the classroom and 8 hours of

on-road driving. The other group (227 students) was not offered high school driver education, leaving them to family or private instruction.

For the 788 participants, a comparison of driving records after licensing was conducted using both self-report information and crash records. Overall, no statistically significant reductions in collisions or convictions were found for AA students. Females in the experimental group had significantly more collisions than those in the control group. Students who were offered high school driver education obtained their licenses earlier, in this case by substantial margins (6 months for males and 10 months for females). The authors suggested that the AA program must be seen as:

... a means to expedite the obtaining of a driver's licence with the advantage of known expert instruction. It should not be regarded as a means to guarantee training in safer driving, nor can one conclude that any habits of safe driving transmitted by this training would outlast the training period. (Wynne-Jones and Hurst 1984, 11)

A similar randomly assigned controlled trial also took place in Australia around the same time, with no significant differences being found in driver records between driver education students and a control group (Strang et al. 1982).

These experimental studies used such small numbers of students that their statistical power was very low. That is, the small number of cases makes it unlikely that even a moderate difference between the crash records of the treatment and control groups would be found to be statistically significant.

Sweden

In 1994, Gregersen reported a modest but elegant Swedish experimental trial that produced surprising results. The educational treatment was a specially developed cooperative program that combined home instruction for the theory component of driver education and coordinated professional in-car instruction (11 hours). Beginner drivers were assigned to either treatment or control groups from among teenagers who reported in a previous survey that they were planning to have only private driving instruction from family or friends. Group assignment was approximately random. The treatment group (about 850) was offered nearly free professional instruction for the trial. Both groups were followed up with surveys over two years.

The educational treatment improved some measures of performance and attitude (reduced reported speed and less overconfidence). The surprising finding was the treatment group was significantly worse in the first year in self-reported crashes per kilometer. In the second year, that group was significantly better in crashes. In looking for an explanation for the unexpected pattern of group differences, Gregersen speculated that cognitive overload might have prevented any benefit of the improved performance appearing during the first year. This suggests that training effects might be more complex than generally assumed, and that careful and prolonged follow-up is essential in a thorough evaluation. These surveys are important, in that they allow

measurement of possible behavioral differences between groups, such as the amount of driving.

California

Dreyer and Janke (1979) conducted an early prospective experimental evaluation in California, and found a substantial benefit in reduced crashes. Structured as a randomized controlled trial, the study compared a range of results for about 2,000 students randomly assigned to driver education programs with and without in-car practice on a special off-road driving range (as opposed to on-road practice only). The total amount of driving time in the two programs was equivalent, but differed in where it took place. The classroom components of the two programs were the same. Unlike the DeKalb study, no attempt was made to include a no-treatment or minimal treatment control condition. Intermediate measures were taken around the time of training, but apparently no attempt was made to acquire data from the students during the follow-up year.

The students who took the assigned range and non-range programs were compared and found to be similar in a number of measures, such as licensing test scores and time to licensing. The non-range students were significantly better in a knowledge test and simulator scores, but the range students were better on a number of driver record measures over the first year of licensed driving. The range students' advantage in total crashes was large (33% lower) and the difference was statistically significant. Other record measures, such as injury crashes and violations were better for range students, but the differences did not reach statistical significance. It is not clear why the range students should have had so many fewer crashes during their first year, since they were not typically better on the limited intermediate criterion measures that were taken. There were no measures during or after the follow-up period, so differences in the amount of driving exposure and other possible differences between the groups are unknown.

A recent study by Masten and Chapman (2003; 2004), also a prospective experimental trial, addressed only intermediate student outcomes using test score results and did not include crash impacts. The study provides a refined modern experimental extension of the historical line of mainly retrospective, quasi-experimental evaluation research that attempted to find differences in outcomes of different forms of driver education and training (see Nichols 2003).

The driver education program being evaluated consisted of a classroom component only, without in-car training. The study's intent was to see whether the diverse delivery methods available and approved in California led to different educational outcomes. These outcomes were measured by scores on specially-constructed, proctored exit tests of knowledge and attitude, as well as the state's licensing knowledge test. The attitude measure was based on the psychological concept "locus of control," which has a modest empirical basis as a correlate or predictor of safe driving (Lonero et al. 1995). A total sample of 1,300 was divided into four different groups for assigning to different training conditions. Use of criteria other than crashes permits this study to use much smaller samples and still retain reasonable statistical power.

The sample of potential new drivers was randomly divided into treatment groups and assigned to one of four instructional settings: classroom, CD-ROM, workbook, or internet/workbook home study. All methods were intended to deliver the same standard driver education curriculum. Group differences on the special exit test and attitude scale were small, but the differences significantly favored the CD-ROM and internet/workbook groups. These methods also had substantially and significantly higher rates of course completion. The DMV knowledge test scores were higher for the classroom group, but this may have been because classroom instructors taught directly to the test items, apparently a common practice.

The Masten and Chapman study is essentially a product evaluation that is narrowly focused on training methods measured against specific intermediate measures of knowledge and attitudes. It is a well-planned and controlled design, however, using reasonably well-tested psychometric measures of intermediate criteria, which in turn have some plausible, although not proven, relation to safe driving.

The study provides useful guidance in developing alternative high-tech delivery of, primarily, the basic knowledge component of beginner driver education. It does not, of course, give direct information about the safety implications of the alternative training delivery methods or of the content delivered. It does, however, raise interesting questions of how such findings should be applied to program development and how they should be followed up with further research. Should the field move toward the more effective methods suggested? Or should those methods be researched further to replicate the findings and further refine the effectiveness of the methods?

Quasi-Experimental Evaluation Studies

United Kingdom

An example of a typical quasi-experimental approach is the U.K. follow-up survey after licensing. In this approach, the relationship between learning to drive and subsequent accidents was measured by a longitudinal three-year survey study of a cohort of newly licensed drivers (Forsyth, Maycock, and Sexton 1995; Maycock 1995; Maycock and Forsyth 1997). About 15,000 new drivers were surveyed by mail questionnaire three times, after one, two, and three years of driving experience. About half returned the first questionnaire with moderate declines in response after the second and third year. Results of this study highlighted differences between males and females. The length of time spent learning to drive, for example, was not related to self-reported accidents for females. For males, however, longer times were associated with fewer subsequent crashes. On average, the U.K. new drivers took about 14 months to learn.

Instruction in this study was limited to in-car lessons. Nearly all respondents had received some professional instruction (males 97%, females 99%). Surprisingly, more instruction was associated with more crashes. In females, where the effect was more clearly significant, the effect seemed to result from a small number of licensing

candidates who: 1) required much instruction before taking the state driving test; 2) were less likely to pass the first time; and 3) crashed more after eventually passing.

Interestingly, only 21% of men and 30% of women reported reading the government driving manual, but this reading was also not significantly related to subsequent crashes. Clearly, this naturalistic type of survey research is limited in its ability to establish causal relations between instruction and crashes due to the self-selected subject population, extraneous, and possibly confounding differences between the groups other than the training received.

Pennsylvania

A more recent similar approach was conducted in Pennsylvania (McKenna et al. 2000). A random sample of 1,188 16- to 18-year-old drivers was selected to be part of a telephone survey that asked respondents to provide information about their driving records and personal demographics. The subjects consisted of high school driver education students (57%), commercial driver education students (13%), and those who reported no formal driver education (34%). Unlike the early quasi-experiments, studies such as this one gathered additional information about the characteristics of the drivers in the driver education and comparison groups, permitting use of multi-variate statistical techniques to partially compensate for the lack of random assignment to the groups.

In the Pennsylvania study, logistic regression was used to simultaneously assess the effects of 16 variables on the outcome of crash or no crash. Crash rates were lower for students with high grades and higher for those who made car payments. No evidence existed of fewer crashes, fewer convictions, increased seat belt use, or lower crash severity for the driver education group.

Manitoba

Manitoba's driver education situation is unusual compared to most other North American jurisdictions. It has a centrally organized high school driver education program supported by Manitoba Public Insurance (MPI), the province's public auto insurer. About 70% of new drivers (14,000) take the course each year.

In 2000-2001, Lonero et al. (2005) conducted a quasi-experimental survey and record study to assess the impacts of the program on students and identify areas for improvement. Drivers' insurance records were used. Driver licensing record data were not available so that some details of licensing and collision information could not be included. Samples consisted of 4,000 each of program graduates and those who learned to drive some other way, as well as 2,000 each of parents and general public. Multivariate statistical methods were used to analyze survey and record data to partially control for pre-existing differences between the driver groups. Mail surveys were used, and questionnaire completions ranged from 26% for the general public to 39% for graduates, considered moderate response rates. The validity of any survey is reduced by the loss of subjects – those who reply probably differ from those who do not in other ways as well. A follow-up brief survey provided data on non-respondents.

After accounting for the age difference between the groups, no significant difference in self-reported crash rates existed between the groups, regardless of whether the comparison group drivers had any formal training. Nearly 20% of young drivers in both groups indicated they had at least one crash not reported to authorities.

Graduates appeared more health conscious and confident in their ability to avoid crashes, and they reported better avoidance of errors and violations. Their greater reported seat belt use showed the strongest difference of any of the self-report items. In recorded convictions, statistically significant differences existed between the groups, but the findings were contradictory. When the whole populations were modeled, the graduates had more convictions in a given year, but when the smaller group of survey respondents was modeled, the comparison group had more violations.

Statistical modeling indicated that age, months of driving experience, and gender were all statistically significant factors in determining crash rate. After controlling for all of these variables, there was no statistically significant crash difference between the groups who had taken driver education or not.

Australia

A recent Australian study used a quasi-experimental survey follow-up of drivers receiving different forms of “pre-driver” education in the state of Victoria (Haworth, Kowadlo and Tingvall 2000). The researchers attempted to compare self-report crash experience of drivers who had taken a pre-driver course including an off-road in-car component, with drivers who had not taken such a course. The comparison group had either taken an in-class-only course or none at all. Multi-variate statistical methods controlled for some of the extraneous differences between the groups (age, sex, and residence location), resulting in an adjusted crash difference of 20% in favor of the drivers with the in-car course.

Because of the small number of subjects (total respondents = 687), however, the difference between the groups was not statistically significant. This seemingly predictable outcome results from a design flaw and is unfortunate in an otherwise rather clever study design. It re-emphasizes the need for careful sample size calculation and power analysis when planning an evaluation study. This study could best be seen as a pilot project.

British Columbia

An ongoing program of evaluation research in British Columbia (Wiggins 2005) is directed primarily to the province’s Graduated Licensing Program (GLP) but also addresses the effects of driver education in the context of the time discount in GLP. In a driving records study that adjusted for age and gender, new drivers who used a driver education certificate to shorten their learner license period in GLP crashed 26% more over their first year of unsupervised driving than those who did not present a certificate. When adjusted for time spent in the learner period, the difference dropped to 13%. A

case-control survey study identified how the new drivers learned to drive. Regardless of whether they used a certificate at licensing, those who took an approved course had about 26% more crashes (adjusted for age, gender, and frequency of driving during the first six months of unsupervised driving).

Wiggins suggests other reasons to believe that driver education in British Columbia may not yet operate at a level consistent with the new GLP standard, but she also points to similar findings appearing in the graduated licensing evaluations in Ontario and Nova Scotia. Factors accounting for the excess risk of driver education graduates in graduated licensing systems are unclear.

Ontario

Zhao et al. (2006) surveyed 1,533 students in numerous Ontario high schools about their driving behavior and related factors, such as class of graduated license held, amount of driving, and crashes. Crash experience was compared for those who had or had not taken driver education, with a number of other factors accounted for by multi-variate statistical models. In this it resembles the Manitoba longitudinal study, although in the Ontario study, self-reported crashes were not supplemented with insurance or licensing records.

Results showed that, among drivers who held Ontario's first stage (learner's) license (G1), significantly fewer driver education graduates reported having crashes than those who had not taken driver education. In fact, among the G1 drivers, driver education was the only factor significantly associated with crashes. In contrast, among drivers with intermediate (G2) licenses, those with driver education had greater odds of reporting a crash, although the difference was not statistically significant. The findings suggest further study is needed to identify reasons for the effect among the highly restricted learner drivers, while no positive effect (and possibly a negative effect) appears for those at the intermediate licensing stage, when they are driving independently.

Texas

A quasi-experimental study of Texas's unique Parent-Taught Driver Education program was conducted. This relatively comprehensive evaluation developed data through focus groups conducted with driver education (DE) instructors, students and parents of DE students; a state-wide mail survey of young drivers; and analysis of driver records. The study attempted to find: 1) differences in the attitudes, knowledge, and skills of novice drivers exposed to the three driving instruction options available in Texas; and 2) differences in post-licensing crash involvement and traffic offense convictions between teenage drivers who are parent taught and those who complete either school based or commercial training.

Based on nine focus groups, a survey of 500 young drivers, and analysis of 1.4 million driver records, the study suggests that the parent-taught driver education program has a negative influence on the overall safety of novice drivers. Few differences in knowledge and skills were found among the modes in self-reports of driving-related knowledge and

skills. However, as measured by State-administered tests parent-taught students demonstrated poorer knowledge early in the training and licensing process and poorer driving skills at the end of training.

During the provisional license period and full licensure, parent taught drivers showed proportionally more total traffic convictions and crashes than the other drivers. The authors recommended that the parent-taught driver education should be improved through better training and monitoring, and that the mandatory government road test should be reinstated.

Ecological Evaluation Studies

The broader scale and less direct ecological approach is the principal alternative to experimental trials and quasi-experiments for evaluating the impact of driver education. In ecological studies, changes or differences in large-scale factors, such as laws or policies, are typically evaluated over time in one jurisdiction or between different jurisdictions.

Connecticut

As a result of eliminating state subsidies for driver education in Connecticut high schools, nine school boards decided to drop the courses, while other communities continued to offer them. Robertson (1980) used this opportunity to investigate the relation between driver education and the age of first licensure and collision rates.

Obtaining a driver's license became more difficult and expensive in areas that dropped driver education courses. Robertson reported that the number of licensed 16- and 17-year-old driver education graduates declined by 57% in the affected communities, compared with 9% in communities where driver education was retained. The affected communities showed a 63% decrease in the collisions of 16- and 17-year-old driver education graduates, while no change was present in the other communities.

With driver education no longer available in the affected communities, declines in licensing and crashes among driver education graduates are not surprising. The total licensing and crashes of all 16- and 17-year-olds in the affected communities declined much less (10-15%). Most young people apparently found other ways to learn to drive. From the data presented, there appear to be slightly fewer crashes per population of 16- and 17-year-olds in favor of the communities that dropped driver education, but only in the second follow-up year. Neither the actual rates for all 16- and 17-year-old drivers, nor any statistical inference regarding the significance of differences, however, are presented in the paper. The study results, therefore, do not strongly support Robertson's conclusions that eliminating driver education improved young driver crash experience significantly by delaying licensure.

Unfortunately, this severely flawed report is still often referenced in support of the suggestion that driver education has significant perverse effects on safety as a result of inducing earlier licensing. Based on their review of DeKalb and other results, Mayhew

and Simpson (1995) concluded that students who take high school driver education are licensed earlier by about one month compared to students who would have taken the course had it been available to them. Also based on DeKalb data, however, they concluded that earlier-licensed driver education students drive less in a given period after licensing, at least partially offsetting the exposure increase that might result from earlier licensure.

Québec

Potvin, Champagne, and Laberge-Nadeau (1988) used a time-series design to evaluate the impact of introducing a mandatory driver-training requirement for all new drivers in Québec. Previously, only new drivers under age 18 were required to be trained. The main effect of the mandatory training requirement was an increase in the total number of crashes, as more 16- and 17-year-old females became licensed, without any reduction in crashes per licensed driver.

Prior to requiring formal training for all new drivers, there may have been a tendency for drivers to avoid the cost of driver training by waiting until the age of 18 to become licensed. The authors theorized that the increase in early licensure occurred because there was no longer any economic advantage to waiting until age 18 to be licensed. The effect was stronger in females, because it was mainly females who had previously waited until after age 18 to become licensed.

U.S. Fatal Crash Modeling Studies

Additional ecological studies in the United States have also failed to find strong beneficial effects of driver education requirements, as concluded in the 1999 review by Vernick and colleagues. Levy (1990), however, conducted a 47-state econometric modeling study of the relationship of various safety factors to fatal crashes of teenaged drivers. He concluded that a mandatory driver education requirement had a small but significant association with fewer fatalities in 15-17 year olds. An earlier modeling study by Robertson and Zador (1978) had failed to find a significant effect of the proportion of driver education graduates on fatalities per 10,000 licensed drivers.

These two studies are unusual in addressing fatal crashes, as nearly all other crash-based evaluations looked at new drivers' total crashes. Fatal crashes are so rare that only these very broad modeling studies, covering large populations of drivers, are able to use them as criteria. Preventing fatal and serious injuries is the main concern of road safety in general, and of young drivers' safety, but nearly all research and evaluation necessarily addresses total crashes, which have been used almost exclusively because of easier availability of large enough numbers of crashes to make moderate differences detectable among the groups being studied. Total crashes consist mainly of minor, property-damage-only crashes, which would not matter if total crashes were a good substitute indicator measure for serious crashes. Theoretical and empirical bases exist, however, for thinking that serious crashes are different from minor crashes.

On generalizing between fatal crashes and all crashes, Robertson and Zador wrote:

Since the characteristics of motor vehicle crashes involving serious injuries are generally similar to those of fatal crashes, it seems likely that these crashes are similarly affected. Run-of-the-mill crashes involving only property damage or minor injuries tend to have quite different characteristics, however, and it is not safe to assume that the conclusions apply to such crashes. (Robertson and Zador 1978, 965)

While any crash study requires large sample sizes, fatal or serious injury crashes could only be considered when huge, national-scale data sets are available. Greater use of large-scale modeling techniques is an important direction for safety R&D and policy support

These approaches are complex and dependent on statistical correction for many factors, but studies like Levy's approach using national data is worth pursuing further to see whether the apparent positive benefit of mandatory driver education provisions in the 1980s can be verified, replicated, or extended in the modern environment.

Denmark

Carstensen's (2002) ecological study examined the effect of a new mandatory driver education program in Denmark consisting of classroom and in-car training that differed from traditional courses by more strongly emphasizing defensive driving and including motorway driving and night driving. Records of 18- and 19-year-old drivers (the youngest eligible to be licensed in the jurisdiction) were followed before and after the new training requirement and compared to older drivers' records. While both age groups' crash rates declined after 1986, the young drivers' crash records declined significantly more.

Driver education can affect subsequent driving exposure, and driving exposure is closely related to crash experience. Unlike many other studies, Carstensen checked for differential changes in driving exposure. Such changes were ruled out as explanations of the new program's greater crash reduction among new drivers.

Other factors in addition to the driver education requirement, however, could have contributed to a decline in young driver crashes. The new training requirement, for example, made becoming licensed more expensive, which may have prevented or delayed licensing for some potential new drivers, perhaps resulting in a higher percentage of older, and lower-risk, new drivers. Specific data on driver age and crashes per driver, however, were not reported. This ecological type of study cannot control for the many potential external forces that could influence the results (Mayhew and Simpson 1997). Adding periodic surveys to supplement record data would add considerable strength to this type of evaluation.

The Danish experience and the later introduction of second-stage training requirements elsewhere in Scandinavia (Keskinen et al. 1998) point to one potentially fruitful direction for further investigation: multi-stage instruction.

Table 1 provides a summary of findings of each of the individual studies discussed previously.

Table 1. Summary of Driver Education Evaluation Results

Reference	Design	Results	Methodological Strengths/Limitations
Experimental Studies			
Dreyer and Janke 1979 <i>California</i>	<ul style="list-style-type: none"> 2,057 students randomly assigned to two training conditions 	<ul style="list-style-type: none"> Those receiving range practice had fewer recorded crashes, but tests scores were no different 	<ul style="list-style-type: none"> Randomized control trial Intermediate measures No follow-up survey for exposure & behavioral measures
Ray et al. 1980 Stock et al. 1983 <i>DeKalb County, Georgia</i>	<ul style="list-style-type: none"> Intensive, minimal, and no driver education groups About 6,000 students randomly assigned to each group 	<ul style="list-style-type: none"> Intensive training (SPC) drivers had better skills and fewer crashes during first 6 months, but not beyond Effects were complex; see text 	<ul style="list-style-type: none"> Comprehensive randomized controlled trial Long follow-up – 6 years Formative evaluations and intermediate outcomes measures
Wynne-Jones and Hurst 1984 <i>New Zealand</i>	<ul style="list-style-type: none"> 788 students, 561 received course, 227 family/friend taught Random assignment 	<ul style="list-style-type: none"> No reduction in collisions for driver education group 	<ul style="list-style-type: none"> Adequate design Small control group No formative evaluation or intermediate outcomes
Gregersen 1994 <i>Sweden</i>	<ul style="list-style-type: none"> 850 students received driver education course compared to controls Random assignment 	<ul style="list-style-type: none"> Driver education group significantly worse first year, significantly better second year 	<ul style="list-style-type: none"> Longer follow-up – 2 years Reasonable sample size
Masten and Chapman 2003; 2004 <i>California</i>	<ul style="list-style-type: none"> 1,300 students randomly assigned to one of four instructional settings 	<ul style="list-style-type: none"> Home-based methods better for 1 knowledge and attitude test, classroom better for DMV knowledge test 	<ul style="list-style-type: none"> Sample size adequate Well planned and controlled Psychometric measures only
Quasi-Experimental Follow-Up Studies			
Forsyth et al. 1995 <i>United Kingdom</i>	<ul style="list-style-type: none"> Survey of 15,000 new drivers 	<ul style="list-style-type: none"> Longer time learning to drive associated with fewer crashes for males More driving education was associated with more crashes 	<ul style="list-style-type: none"> Several follow-ups over time Self-selection bias Self-reported data only

Reference	Design	Results	Methodological Strengths/Limitations
Howarth et al. 2000 <i>Australia</i>	<ul style="list-style-type: none"> • Self-report crash effects for in-car training effects 	<ul style="list-style-type: none"> • Substantial differences, but not significant 	<ul style="list-style-type: none"> • Sample size too small
McKenna et al. 2000 <i>Pennsylvania</i>	<ul style="list-style-type: none"> • Survey and crash records • Random sampling for survey 	<ul style="list-style-type: none"> • Driver education not associated with lower crashes or convictions 	<ul style="list-style-type: none"> • Multi-variate statistical analysis used to control for confounding variables • SES missing from control variables
Lonero et al. 2005 <i>Manitoba</i>	<ul style="list-style-type: none"> • Survey and crash records • Random sampling for survey 	<ul style="list-style-type: none"> • Driver education not associated with lower crashes or convictions 	<ul style="list-style-type: none"> • Multi-variate statistical analysis used to control for confounding variables
Wiggins 2005 <i>British Columbia</i>	<ul style="list-style-type: none"> • Cohort record study • Case control study with survey and records 	<ul style="list-style-type: none"> • New graduated license holders who took driver education had 26% more crashes 	<ul style="list-style-type: none"> • Multi-variate statistical analysis used to control for confounding variables
Zhao et al. 2005 <i>Ontario</i>	<ul style="list-style-type: none"> • Self-report survey of high school students 	<ul style="list-style-type: none"> • Driver education associated with fewer crashes for learner license holders 	<ul style="list-style-type: none"> • Multi-variate statistical analysis used to control for confounding variables
Pezoldt et al. (2007) <i>Texas</i>	<ul style="list-style-type: none"> • Focus groups, surveys, and driver records 	<ul style="list-style-type: none"> • Parent-taught teens less safe. 	<ul style="list-style-type: none"> • Comprehensive approach with intermediate measures
Ecological Studies			
Robertson and Zador 1978 <i>27 U.S. States</i>	<ul style="list-style-type: none"> • Modeling study of driver education and fatal crash rates 	<ul style="list-style-type: none"> • No relation between proportion taking driver education and fatality rates 	<ul style="list-style-type: none"> • Not program specific
Robertson 1980 <i>Connecticut</i>	<ul style="list-style-type: none"> • School boards with and without driver education 	<ul style="list-style-type: none"> • For school boards without driver education, total licensing and crashes of 16- and 17-year-olds decreased by 10-15% 	<ul style="list-style-type: none"> • Not enough data analysis presented
Potvin et al. 1988 <i>Québec</i>	<ul style="list-style-type: none"> • Mandatory driver education introduced in Québec for all (formerly just 16-17 year olds) 	<ul style="list-style-type: none"> • Increased number of young driver crashes due to increased number of licensed females aged 16-17 	<ul style="list-style-type: none"> • Large sample size • Different timeframes for treatment and control groups
Levy 1990 <i>USA 47 States</i>	<ul style="list-style-type: none"> • Large-scale modeling study of effects of mandatory driver education 	<ul style="list-style-type: none"> • Small but significant beneficial effect on fatal crashes 	<ul style="list-style-type: none"> • Not program specific

Reference	Design	Results	Methodological Strengths/Limitations
Carstensen 2002 <i>Denmark</i>	<ul style="list-style-type: none"> Mandatory driver education, new curriculum 	<ul style="list-style-type: none"> Reduced crashes 	<ul style="list-style-type: none"> Large sample size No control of confounding variables

IMPLICATIONS FOR DRIVER EDUCATION EVALUATION

A major concern for all kinds of evaluation is the criterion of effectiveness. What *exactly* do we want and expect the program to achieve? Traditional criterion for effectiveness of driver education has been to reduce crashes among graduates relative to those who learned to drive in other ways. When considering safety effects, the specific measure of crash experience is important and can be controversial. Different ways of measuring crashes can provide quite different results. If crashes are measured through self-report, government records, or insurance records, differences arise due to completeness of reporting and different timing for records to work their way through bureaucratic systems.

Crash experience is best reported in terms of rates, and both the numerator and the denominator chosen for the rate have important implications. Crashes of different severity, persons injured, and persons killed as numerators can also provide different perspectives. A key distinction among crashes as a criterion is what kind of crashes, but this has rarely been discussed in driver education literature. Clearly, the concern with young driver safety is for serious injury crashes, particularly those leading to death and permanent disability. Property-damage crashes are less of a concern, except from an insurance or cost perspective. However, crash-based analyses almost always count total crashes, the great majority of which are minor property damage crashes. Since it seems clear that fatal crashes are often different than minor crashes in terms of circumstances and etiology, this is a fundamental weakness in evaluation.

As seen in the various DeKalb analyses, the rate denominator is also important. Crashes per licensed driver can give quite different results than crashes per assigned experimental subject. Crashes per distance driven can give different results from crashes per person, as seen in the Elvik and Va (2004) meta-analysis. The rate denominator that is selected needs to match the goals of the program and the evaluation. Should the success criterion for driver education be safer mobility, or the broader public health goal of a safer population? Preference for mobility-based rates, such as crashes per mile traveled versus preference for population-based measures, such as crashes per age-group population, reflect fundamental theoretical differences and need to be resolved early in evaluation planning.

Ultimate safety measures are important success indicators, but they are not the only important educational objective for driver education, or any form of safety education. This is particularly clear where the safety education is sequenced and coordinated with other influences, which is a condition now thought to be critical to success. Donelson

and Mayhew's (1987) extensive review of driver improvement programs emphasized that concentrating on intermediate outcomes was critical to more effective driver improvement interventions. These, like driver education, are usually intended to have educationally-driven behavior change effects on crash propensity. In either case, focusing only on the typical lack of significant safety effects makes it unlikely that programs can be improved enough to actually achieve the desired safety effects, because such evaluations do not point toward ways of improving programs.

In the mid-1980s, the Road Transport Research Programme of the Organization for Economic Cooperation and Development (OECD) assessed the efficacy of road safety education programs and provided strategies for program development, implementation, and evaluation (OECD 1986). This report outlined issues of effectiveness for program planning and implementation, but its primary focus was evaluation. It was suggested that a program should be seen as effective if it does what it was intended to do, and that it is, therefore, very important to be explicit about educational objectives, which should include measures of intermediate effects and not be solely focused on collisions.

Clear objectives addressing drivers' task requirements are crucial to program effectiveness, but are problematical because of the lack of empirical knowledge available on many driver tasks. Intermediate objectives, such as better knowledge, skills, and attitudes are currently difficult because of lack of research to link them to collisions, and research should be undertaken to validate such measures.

The content and structure of instruction were also addressed by the 1986 OECD report. Internal consistency between content and objectives is critical, but little was then known about the relationship between structures of instruction (where, how much, how often, etc.) and effectiveness. With respect to the process of instruction, three sets of critical variables were identified: learner variables, instructional variables, and social variables. Attitudes and a host of other learner factors must be considered when designing an effective program. Instructional variables are somewhat easier to clarify. Instructors must be skilled and motivated, and the report suggested that teachers are usually not well qualified to teach safety education on either of these criteria. Recent research has started to meet this need.

Critical social variables included socioeconomic and cultural differences, and language. The OECD report suggested that acceptance of driver education and its importance by teachers and other potential delivery agents is so difficult in many cases, particularly in secondary schools, that it would be better to train special staff to instruct through broadcast media and closed TV networks. While this may have seemed a far-fetched and gloomy idea at the time it was made, recent trends toward computer- and web-based driver instruction could be seen as validating the OECD prognostication of driver education going directly to the student outside school settings.

The OECD committee saw evaluation as the key to successful safety education. The report identified three evaluation levels. To structure the OECD model, the committee used the *formative* and *summative* evaluation concepts originally suggested by the eminent evaluation theorist Michael Scriven (Stufflebeam 2001).

In the OECD model, two types of formative level evaluation were identified: 1) process evaluation – how a program is used and received; and 2) product evaluation – impacts on skills, knowledge, attitudes, or behavior. The third type was summative or outcome evaluation, featuring two kinds of measures – cost/benefit, and the driver education program fitting with the education system at large.

Evaluation is especially important in education programs, where the interventions seem as if they ought to be effective and are so obviously desirable. Aside from lack of positive effects, the possibility of negative effects was recognized by the OECD, which pointed out that some kinds of advanced skills training can make some drivers less safe (OECD Scientific Expert Group 1990). This makes both formative and summative evaluation especially critical for education programs. Any potentially effective behavioral technology may be ineffective, or even harmful, depending on how it is applied.

A related weakness is inherent in quasi-experimental studies, where attempts to make clean, unbiased comparisons involve multi-variate statistical methods to partially control for extraneous factors that might bias the comparison. Identifying and controlling all likely biases in the characteristics of non-randomly assigned groups are difficult, and unlikely to be perfect in any one study. The evaluation of Pennsylvania's driver education program (McKenna et al. 2000), for example, identified 16 control variables but ignored socioeconomic status, typically an important factor in young drivers' risk differences.

With notable exceptions, such as the DeKalb experiment, Dreyer and Janke's 1979 experimental study, and Gregersen's 1994 survey study, most evaluations have failed to look at intermediate measures. Lacking information on what the students have or have not learned, directions for program improvement are left unclear. Most existing evaluations leave many unanswered questions regarding:

- Logical links between curricula and young drivers' needs;
- Theories explaining how a program is expected to achieve safer driving;
- The quality, comprehensibility, and usability of curriculum products;
- How well and how consistently the instructional processes actually deliver the intended learning; and
- Which learning and behavioral outcomes result, or fail to result, from the training.

Tracking of learning outcomes is an area where programs could easily build in ongoing evaluation. Knowledge and attitude measures at the beginning and end of the course and at later intervals would help keep contact with graduates and provide feedback for continuous improvement of curriculum and delivery.

The Dreyer and Janke (1979) study shows that intermediate measures may not be enough to help explain crash results. In that case, the few differences found in intermediate measures seemed to favor the group that subsequently had *more* crashes. This study did not include surveying the new drivers during the follow-up period, however, so possible differences in driving behavior or in amount of driving were not

measured. Such information is critical in explaining how an apparently effective training program actually has its effects.

Further, as reflected in Dreyer and Janke's study, exposure differences are too often ignored in driver education evaluation. In general, simple differences in the amount of driving (exposure to risk) account for much of the differences in crash rates between groups of drivers (Chipman 1982). Apart from the simple amount of driving, qualitative exposure differences, such as time of day, presence of passengers, geographic areas, and different trip purposes also represent different levels of collision risk, especially for young drivers (Preusser 1996; Preusser, Ferguson, and Williams 1997). Since different methods of learning to drive may be related to when the beginner chooses to be first licensed and subsequently exposed to risk, exposure information is important in any attempt to evaluate driver education programs. Otherwise exposure could be an important source of confounding or bias in the evaluation results.

The relative scarcity of evaluation work in driver education following the OECD report is unfortunate. As indicated earlier, a serious lack of statistical and research design sophistication was evident in the early studies. The later studies used progressively better designs and statistical methods. No perfect study exists. As discussed in the following section, even the random controlled trials have had weakness in terms of maintaining clean assignment of students to training conditions and clear comparisons between assigned groups.

Randomized Controlled Trials: The Comparison Problem

Some of the liveliest discussions during the development of the *Guidelines* involved using experimental designs in evaluation – particularly RCTs. In this powerful research method, people are randomly assigned to either a treatment group the members of which some specific treatment, or to a control group, the members of which do not receive treatment. This allows researchers to conclude with a reasonable and calculable level of confidence that any between-group differences in outcomes are actually caused by the different treatments received. Since the groups are assigned randomly, other differences among the people should be distributed more or less evenly between the two groups.

This way, RCTs reduce the effects of confounding factors, such as differences between the groups, aside from the actual treatments received, that might cause different results. These could include, for example, differences between people who choose to take driver education and those who do not. In non-experimental designs, one cannot be sure that such confounding differences between the groups have not influenced the results. Even with random assignment, differences may still arise by chance, but statistical methods can calculate a good estimate of the probability that the observed differences could have resulted from chance.

Because of the “clean” comparison RCT permits, it is considered the “gold standard” for establishing causal relationships in the experimental sciences (Williams et al, 2009). Indeed the RCT experimental paradigm is held in such high esteem that non-RCT

studies are sometimes regarded as non-scientific and are disregarded. In Roberts and Kwan's *Cochrane Library* systematic review of driver education (2004), for example, only RCT studies were considered. On the other hand, education evaluators are less favorable to experimental methods and RCTs.

What do these different views for the driver education evaluation imply? Can we simply accept that RCTs are more scientific than all other methods? Since experimental methods are hardly used at all in some sciences, such as astrophysics, economics, and epidemiology, this simple scientific/non-scientific distinction seems unsupportable. Although RCTs are good ways to structure many kinds of research, they are not the only scientific ways.

In evaluating beginner driver education, basic practical problems surface with RCTs. To see this clearly, we need to look in detail at RCTs applied to beginner driver education. Numerous evaluations in the field have been RCTs, including the benchmark DeKalb study (Stock et al. 1982) and the roughly contemporary California study (Dreyer and Janke 1979). Both of these evaluations randomly assigned subjects who wanted to take driver education to receive some form of treatment. The DeKalb study assigned subjects either to one of two groups receiving different high school driver education programs or to a third ("no-treatment" control group), which was not supposed to receive formal driver education. The California RCT assigned subjects to two different training formats (training on a closed-course driving range vs. on-road). A no-treatment control group was not included.

In beginner driver education evaluation, the RCT paradigm runs into difficulty when choosing a comparison condition against which to assess the results of the target program. Indeed, thoughtful researchers have suggested that, for evaluation of beginner driver education, the most suitable comparison is with informal driver training by parents. Comparing driver education to no training at all isn't possible, of course, since all new drivers must somehow learn the basic driving skills.

Assigning students who apply for driver education to even an informal training control group is difficult in practical terms. As DeKalb showed, effectively denying formal driver education to students who want it is not easy, at least in part because it may mean students have to forego insurance premium discounts. To work around the problem in DeKalb, the usual insurance discount was offered to the informal training control students if they passed a special road test. This probably compromised the control group, but it is not known how the group prepared for taking the special road test or how many prepared and took the test.

The difficulty in creating and maintaining a clean control group comparison seems to be an inevitable problem with evaluating beginner driver education using an RCT design. It should not be seen as an avoidable error in the DeKalb study, although better tracking of students through repeated surveying could have helped the analyses by permitting statistical control for some of the difficulties in maintaining clean group assignment.

If comparison with no formal training is viewed as absolutely necessary, then the closest approximation would be complex quasi-experimental designs, which try to statistically compensate for confounding differences between the groups. This conundrum seems unique to evaluating *beginner* driver education. Other forms of driver instruction, such as driver improvement programs, can be withheld from a randomly assigned no-treatment group.

Are there alternative approaches that get around the comparison problem in RCTs? The California RCT study simply compared two different forms of driver education, without an informal training control group (Dreyer and Janke 1979). Assuming the two programs are not so different that they provoke switching groups, this can provide an opportunity for clean random assignment and maintenance of equivalent groups. Unfortunately, most reviewers have ignored Dreyer and Janke's study. It is not clear why, but it may be in part because of the lack of a no-treatment control group (Peck, personal communication). This leads to a potential problem of interpretation – uncertainty of whether or not the “better” treatment had a real positive impact. It is possible that the apparently better treatment had no effect, whereas the “worse” treatment yielded a negative effect, or even that the “better” treatment was actually the less detrimental of two treatments that both had negative safety impacts. Although these possibilities may seem remote, they cannot be casually dismissed, as some driver education evaluations have found negative impacts on some measures.

A study in Sweden found another way around the comparison problem. It selected its comparison groups from young people who were *not* planning to take formal driver education (Gregersen 1994). Some students were assigned to a formal training condition and induced to take the training offered. This could result in a clean comparison with informal training, since the comparison group will probably not seek training. This approach, however, leads to a problem of generalizability. Some might argue that the results of such an evaluation do not necessarily apply to most new drivers whose inclination is to take driver education.

A key implication of these difficulties is that there is no perfect method for evaluation of beginner driver education. RCTs can be useful, but they should have a broader and more systematic evaluation approach, which includes other study designs, a wide range of output and outcome measures, and a variety of carefully planned comparisons. Perhaps it is time the “gold standard” title is passed along to comprehensive, systematic evaluation that includes various approaches to developing the whole picture of program effects and improving programs.

The final implication of past evaluations of driver education is how little they have contributed to developing and improving the programs. Earlier thinking, including that behind the DeKalb project, seems to reflect the idea that a single massive development effort could achieve driver education's safety goals in one step. More recently, however, researchers and theorists have emphasized the importance of incrementally building knowledge gains and other systematic intermediate effects, (Lonerio et al. 1994; Woolley 2000), as well as continuously developing and improving programs (Keskinen et al. 1998). Keskinen et al. wrote: “We have decided ... that the development of driver

education will take place in short steps, with constant evaluations of the results and trying to avoid solutions which are thought to be final” (p. 382).

DISCUSSION

Driver education is intended to reduce crashes and, more particularly, injuries and fatalities. While fatal and other serious crashes are the main concerns of road safety, driver education has almost always been evaluated on total crash rates, which consist mostly of minor crashes. Crashes are rare events with complex causation. Minor crashes mainly result from simple, inadvertent errors and may be hard for novice drivers to control. Fatal and serious-injury crashes are much rarer, and they have different patterns of causation, are often related to serious misbehavior, and they may actually be more amenable to intentional control by the driver.

If driver education is to achieve its safety goals, it probably needs to be more firmly based on research and theory concerning driver skills, behavior, motivation, and risk, and the best ways of influencing them. Program evaluation is critical for more effective program development; past evaluation weaknesses must be recognized so they can be corrected and ongoing evaluation can contribute to improvement in the ultimate criterion of success – rates of serious crashes. A more comprehensive approach to evaluation is needed to address theory, products, processes, program management, and intermediate student outcomes. Even more effective evaluation of safety outcomes will not necessarily point toward how to achieve improved programs. To support this need, formative evaluation is a critical part of comprehensive evaluation.

Evaluating driver education effects on intermediate outcome criteria should include changes in behavior, knowledge, attitudes, and exposure to risk. Intermediate measures should continue during the follow-up period if we are to have a clear picture of the effects of the program and the reasons for them. Reflecting these needs, the Large Scale Evaluation of Driver Education (LSEDE) research program includes a wide range of intermediate measures of driver knowledge, skills, attitudes, and behaviors.

Meeting the ultimate goal of reducing novice drivers' serious crashes will also likely require evaluating and managing the context of driver education. This would involve assessing the linkage of driver education with parental and community influences, graduated licensing, and other behavioral influences, such as incentives and disincentives. The NHTSA sponsored National Driver Education Administrative Standards call for better integration of driver education with graduated licensing, and the impacts of the standards should be evaluated.

The now aging DeKalb study has been considered to be the most extensive and rigorous driver education evaluation, but even this study had serious limitations, and its conclusions are still controversial. Most other evaluation studies were more limited in scope and scale. Experimental evaluations typically have found no statistically significant effects of driver education on crash records, but some analyses of the DeKalb data and one California study did show positive effects. Several quasi-

experimental and ecological studies have been conducted. Two large-scale ecological evaluations showed positive effects of driver education, but one early study did not. No one study design is perfect, and progress will likely develop on a “weight-of-evidence” basis over numerous studies of different types.

A main concern with driver education in schools has been that it leads to earlier licensing and may lead to increased driving exposure. Earlier licensing of new drivers may occur for numerous reasons. Some of these reasons are related to driver education, for example, when parents consider their children to be well trained and, therefore, safe to drive. The effect of taking driver education on earlier licensing seems to be fairly consistent but more modest in the amount of time of earlier licensing than some researchers have feared.

Aside from earlier licensing leading to more exposure, better-trained students may become overconfident, and this may offset the superior skills and knowledge. Relatively new research reinforces the idea that some forms of driver education and training can lead to overconfidence. These considerations lead to questions of suitable goals and expectations for driver education – whether it is supposed to support safe mobility, to enhance public health, or both. Such considerations have important implications for the logic of driver education evaluation and program development.

Curriculum content and structure, standards and governance, market incentives, and integration with licensing are developing rapidly in driver education. However promising these developments may be, evaluation still seems to be more the exception than the rule, and whether the new developments are more effective than traditional approaches is not clear. Real progress in beginner driver education requires more systematic and comprehensive evaluation.

It is important to recognize the limitations of past evaluations. This recognition has been lacking in the past and has led to unfortunate policy decisions, particularly in the 1980s, cutting of R&D and other support. Driver education evaluation, like driver education itself, is evolving and still far from its ideal state.

SUMMARY AND CONCLUSIONS

Past studies have demonstrated that common-sense assumptions about what is effective in reducing young driver crash risk are not always well founded. Reviewers of the evaluation literature have typically concluded that beginner driver education has yet to demonstrate clear success in improving safety of new drivers. A few studies have shown positive effects, but these have been ignored by most reviewers and have not been followed up by either the research community or the program community.

When looking critically at past evaluations, one is struck by how little they have contributed to developing and improving driver education. Evaluation research in driver education has been unsystematic, in the sense that studies typically failed to build on earlier research. Surprising negative findings, and even credible positive findings, were not followed up by further research. Scientific knowledge usually develops through

systematic replication of research, but that has not yet happened in driver education evaluation research. Earlier thinking, including that behind the DeKalb project, seemed to favor a single massive development effort to achieve driver education safety goals in one big step. More recently, however, researchers and theorists have emphasized the importance of incrementally building knowledge, skills, attitudes, and motivations, as well as continuously developing and improving programs.

Certainly there are perceptible trends and renewed development in most aspects of this increasingly diverse field. A renewal of federal leadership in the U.S., embodied in the new National Administrative Standards, and central direction in other countries bodes well for the future. However, most of the factors that have constrained healthy development in the past are still in place.

Theory in driver education is still weak and shows little improvement. Driver education delivery is highly fragmented, and both consolidation and further fragmentation appear to be taking place simultaneously. Driver education needs to be more firmly based in sound research and theory concerning young drivers and, at the same time, in the principles of effective behavior change. It needs better management of the linkage of driver education with parental and community influences, graduated licensing, and other behavioral influences such as incentives and cultural factors.

Comprehensive and systematic evaluation research can be a constructive and important part of future development in driver education. A comprehensive approach to evaluation addresses program theory, context, products, processes, and management, as well as outcomes and impacts. The need for such research is increasing, as vigorous development is occurring in some public and private programs. The previously published AAAFTS Guidelines have provided new materials and direction, and support for more systematic evaluation appears to be growing. If the apparent trends toward data-driven development can be sustained and expanded, they could ultimately lead to improved safety outcomes.

REFERENCES

- Boyce, T. E., and E. S. Geller. 2002. An instrumented vehicle assessment of problem behavior and driving style: do younger males really take more risks? *Accident Analysis and Prevention* 34 (1):51-64.
- Carstensen, G. 1994. *Evaluation of a new driver education in Denmark. Linköping, Sweden*: Swedish Road Transport Research Institute.
- Carstensen, G. 2002. The effect on accident risk of a change in driver education in Denmark. *Accident Analysis and Prevention* 34 (1):111-121.
- Centers for Disease Control and Prevention. 1999. Framework for program evaluation in public health. *MMWR*. 48 (No.RR-11):1-42.
- Chipman, M. L. 1982. The role of exposure, experience and demerit point levels in the risk of collision. *Accident Analysis and Prevention* 14 (6):475-483.
- Christie, R. 2001. *The effectiveness of driver training as a road safety measure: A review of the literature*. Melbourne, Australia: Royal Automobile Club of Victoria (RACV) Ltd., Public Policy Group.
- Christie, R. 2002. Road safety education and training from a public health perspective. *Road Safety Research, Policing and Education Conference Proceedings*.
- Clinton, K. & Lonerio, L. (2006) *Evaluation of Driver Education: Comprehensive Guidelines*, Washington, D.C.: AAA Foundation for Traffic Safety.
- CPBH The Center for the Advancement of Community Based Public Health. 2000. *An evaluation framework for community health programs*. Durham, NC: CPBH.
- Davis, C. S. 1990. *The DeKalb County, Georgia, Driver Education Demonstration Project: Analysis of its long-term effect*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Donelson, A. C., and D. R. Mayhew. 1987. Driver improvement as post-licensing control: The state of knowledge. In *Nonpunitive approaches to driver improvement: Rewards, incentives, and reinforcers*. Toronto, ON: Ontario Ministry of Transportation and Communications.
- Dreyer, D., and M. Janke. 1979. The effects of range versus nonrange driver training on the accident and conviction frequencies of young drivers. *Accident Analysis and Prevention* 11:179-198.

- DSAA (Driving School of the Americas) and RSEA (Road Safety Educators' Association). 2005. *Process for curriculum review*. Ontario, Canada: DSAA and RSEA.
- Elvik, R., and T. Vaa. 2004. *The handbook of road safety measures*. Elsevier.
- Engström, I., N. P. Gregersen, K. Hernetkoski, E. Keskinen, and A. Nyberg. 2003. *Young novice drivers, driver education and training: Literature review*. Linköping, Sweden: Swedish National Road and Transport Research Institute.
- Evans, L. 1987. Young driver involvement in severe car crashes. *Alcohol, Drugs and Driving* 3 (3-4):63-78.
- Evans, L. 1991. *Traffic safety and the driver*. New York: Van Nostrand Reinhold.
- Fisher, D. L., N. E. Laurie, R. Glaser, K. Connerney, A. Pollatsek, S. A. Duffy, and J. Brock. 2002. Use of a fixed-base driving simulator to evaluate the effects of experience and PC-based risk awareness training on drivers' decisions. *Hum Factors* 44 (2):287-302.
- Forsyth, E., G. Maycock, and B. Sexton. 1995. *Cohort study of learner and novice drivers: Part 3, Accidents, offences and driving experience in the first three years of driving*. Crowthorne, Berkshire: Transport Research Laboratory.
- Fuller, R., and D. Bonney. 2003. Driver education and training in post-primary schools in Ireland. In *Behavioural research in road safety: Thirteenth seminar*. Crowthorne, Berkshire: Transport Research Laboratory.
- Fuller, R., and D. Bonney. 2004. *Driver education and training in post-primary schools. Monograph*. Dublin: National Council for Curriculum and Assessment.
- Furby, L., and R. Beyth-Marom. 1990. *Risk taking in adolescence: A decision-making perspective*. Washington, DC: Carnegie Council on Adolescent Development.
- Gebers, M. A., Peck, R. C., Janke, M. K., & Hagge, R. A. (1993). Using traffic violator school citation dismissals in addition to convictions as the basis for applying postlicense control actions. Sacramento, CA: California Department of Motor Vehicles.
- Glad, Alf. 1988. *Phase 2 driver education: Effect on the risk of accident*. Oslo, Norway: Norwegian Center for Research, Department of Transportation.
- Goldman, K. D., and K. J. Schmalz. 2001. Theoretically speaking: Overview and summary of key health education theories. *Health Promotion Practice* 2 (4):277-281.

- Gonzales, M. M., L. M. Dickinson, C. DiGuseppi, and S. R. Lowenstein. 2005. Student drivers: A study of fatal motor vehicle crashes involving 16-year-old drivers. *Annals of Emergency Medicine* 45 (2):140-146.
- Green, L. W., and M. W. Kreuter. 1999. *Health promotion planning: An educational and environmental approach*. 3rd ed. Mountain View, CA: Mayfield Publishing (McGraw-Hill).
- Gregersen, N. P. 1994. Systematic cooperation between driving schools and parents in driver education, An experiment. *Accident Analysis and Prevention* 26 (4):453-461.
- Haworth, N., N. Kowadlo, and C. Tingvall. 2000. *Evaluation of pre-driver education program*. Victoria, Australia: Monash University Accident Research Centre.
- Highway Safety Center. 2002. *Traffic safety education life long learning process*. Indiana, PA: Indiana University of Pennsylvania.
- Irwin, C. E., Jr., and S. G. Millstein. 1992. Risk-taking behaviors and biopsychosocial development during adolescence. In *Emotion, cognition, health, and development in children and adolescents*, edited by E. J. Susman, L. V. Feagans and W. J. Ray. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Joint Committee on Standards for Educational Evaluation. 1994. *The Program Evaluation Standards: How to assess evaluations of educational programs*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- Katila, A., E. Keskinen, S. Laapotti, and M. Hatakka. 1995. *Changes in slippery road accidents as an effect of renewed driver training in Finland*. Turku, Finland: University of Turku.
- Keskinen, E., M. Hatakka, and A. Katila. 1998. *The Finnish way of driver improvement: Second phase of training*. Paper presented at International Driver Improvement Workshop, at Bergisch Gladbach, Germany: Bundesanstalt fur Strassenwesen.
- Levy, D. T. 1988. The effects of driving age, driver education, and curfew laws on traffic fatalities of 15-17 year olds. *Risk Analysis* 8 (4):569-574
- Levy, D. T. 1990. Youth and traffic safety: The effects of driving age, experience, and education. *Accident Analysis and Prevention* 22 (4):327-334.
- Lonero, L., and K. Clinton. 1998. *Changing road user behavior: What works and what doesn't*. PDE Publications. Available from www.drivers.com.
- Lonero, L., K. Clinton, J. Brock, G. Wilde, I. Laurie, and D. Black. 1995. *Novice driver education model curriculum outline*. Washington, DC: AAA Foundation for Traffic Safety.

- Lonero, L. P., K. M. Clinton, and D. Black. 2000. *Training to improve the decision making of young novice drivers. Volume II: Literature review: Consistency not capacity: Cognitive, motivational and developmental aspects of decision making in young novice drivers*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Lonero, L. P., K. M. Clinton, G. J. S. Wilde, K. Roach, A. J. McKnight, H. Maclean, S. J. Guastello, and R. Lamble. 1994. *The roles of legislation, education, and reinforcement in changing road user behaviour*. Toronto, Ontario: Ontario Ministry of Transportation.
- Lonero, L. P., K. M. Clinton, B. N. Persaud, M. L. Chipman, and A. M. Smiley. 2005. *A longitudinal analysis of Manitoba Public Insurance Driver Education Program*. Winnipeg, Manitoba: Manitoba Public Insurance.
- Lonero, L., Clinton, K., Sleet, D. (2006). Behavior Change Interventions in Road Safety.(in) Geilen, A.C., Sleet, D.A., & DiClemente, R.J. *Injury and Violence Prevention: Behavioral Science Theories, Methods, and Applications*, San Francisco, California: Jossey-Bass.
- Lonero, L. (2008) Trends in driver education evaluation and development. *American Journal of Preventive Medicine*;35 (3S):S316–S323
- Lund, A. K., A. F. Williams, and P. Zador. 1986. High school driver education: Further evaluation of the DeKalb County study. *Accident Analysis and Prevention* 18 (4):349-357.
- Masten, S. V., and E. A. Chapman. 2003. *The effectiveness of home-study driver education compared to classroom instruction: The impact on student knowledge, skills, and attitudes. Report to the legislature of the state of California*. Report no. CAL-DMV-RSS-03-203. Sacramento, CA: California Department of Motor Vehicles.
- Masten, S. V., and E. A. Chapman. 2004. The effectiveness of home-study driver education compared to classroom instruction: The impact on student knowledge and attitudes. *Traffic Injury Prevention* 5 (2):117-121.
- Maycock, G. 1995. Accidents in the first three years of driving. *TRL Annual Review*: 71-77.
- Maycock, G., and E. Forsyth. 1997. *Cohort study of learner and novice drivers. Part 4: Novice driver accidents in relation to methods of learning to drive, performance in the driving test and self assessed driving ability and behaviour*. Crowthorne, Berkshire: Transport Research Laboratory.

- Mayhew, D. R., and H. M. Simpson. 1995. *The role of driving experience: Implications for the training and licensing of new drivers*. Toronto, ON: Insurance Bureau of Canada.
- Mayhew, D.R., and H.M. Simpson. 1997. *Effectiveness and role of driver education and training in a graduated licensing system*. Ottawa, Ontario: Traffic Injury Research Foundation.
- Mayhew, D. R., H. M. Simpson, and A. Robinson. 2002. The safety value of driver education and training. *Injury Prevention* 8 (Supplement II):3ii-8.
- Mayhew, D. R., H. M. Simpson, and A. Pak. 2003. Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis and Prevention* 35 (5):683-691.
- McCartt, A.T., Mayhew, D.R., Braitman, K.A., Ferguson, S.A., and Simpson, H.M. (2008) *Effects of Age and Experience on Young Driver Crashes: Review of Recent Literature*. Insurance Institute for Highway Safety, Arlington, Va.
- McKenna, C. K., B. Yost, R. F. Munzenrider, and M. L. Young. 2000. *An evaluation of driver education in Pennsylvania*. Harrisburg, PA: Pennsylvania Department of Transportation.
- McKnight, A. J. 1984. Driver education—when? *Journal of Traffic Safety Education* (Jan):13-15.
- McKnight, A. J., and A. S. McKnight. 2003. Young novice drivers: Careless or clueless? *Accident Analysis and Prevention* 35 (6):921-925.
- National Driver Education Standards Project. 2009. *Novice Teen Driver Education and Training Administrative Standards*. US Department of Transportation -- National Highway Traffic Safety Administration, Washington, D.C.
- Nichols, J. L. 2003. *Driver education review*. Paper presented at the NTSB Forum on Driver Education and Training, October, 2003. Washington, DC: National Transportation Safety Board.
- NIMH (National Institute of Mental Health). 2004. *Teenage brain: A work in progress*. NIMH, 05/13/04 2001 (accessed November 16, 2004). Available from <http://www.nimh.nih.gov/publicat/teenbrain.cfm>.
- NTSB (National Transportation Safety Board). 2005. *National Transportation Safety Board public forum on driver education and training, October 29-30, 2003. Report of proceedings*. Washington, DC: NTSB.
- OECD. 1986. *Effectiveness of road safety education programmes*. Paris: Road Transport Research, Organisation for Economic Co-operation and Development.

- OECD Scientific Expert Group. 1990. *Behavioural adaptations to changes in the road transport system*. Paris: Road Transport Research, Organisation for Economic Co-operation and Development.
- Patton, M. Q. 1997. *Utilization-focused evaluation: The new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications.
- Peck, R. C., & Gebers, M. A. (1991). The traffic safety impact of TVS citation dismissals (Report # 133). Sacramento, CA: California Department of Motor Vehicles.
- Pezoldt, V. J., Womack, K. N., and Morris, D. E. (2007). *Parent-Taught Driver Education in Texas: A Comparative Evaluation*. U.S. Department of Transportation, National Highway Traffic Safety Administration, DOT HS 810 760.
- Plato, K. C., and A. Rasp, Jr. 1983. *Parental views on issues related to traffic safety education and the licensing of teenage drivers*. Olympia, Washington: Washington Office of the State Superintendent of Public Instruction.
- Potvin, L., F. Champagne, and C. Laberge Nadeau. 1988. Mandatory driver training and road safety: The Quebec experience. *American Journal of Public Health* 78:1206-1209.
- Preusser, D. F. 1996. Licensing practices and crash risk in the United States. In *New to the road: Reducing risks for young motorists. Proceedings of the First Annual Symposium of the Youth Enhancement Service, June 8-11, 1995*, edited by H. Simpson. Los Angeles, CA: Youth Enhancement Services, University of California.
- Preusser, D. F., S. A. Ferguson, and A. F. Williams. 1997. *Effect of teenage passengers on the fatal crash risk of teenage drivers*. Arlington, VA: Insurance Institute for Highway Safety.
- Quadrel, M. J., B. Fischhoff, and W. Davis. 1993. Adolescent (in)vulnerability. *American Psychologist* 48 (2):102-116.
- Ray, H. W., M. Sadof, J. Weaver, J. R. Brink, and J. R. Stock. 1980. *Safe performance secondary school driver education curriculum demonstration project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Roberts, I., and I. Kwan. 2004. School based driver education for the prevention of traffic crashes. *The Cochrane Library* (1).
- Robertson, L. S. 1980. Crash involvement of teenaged drivers when driver education is eliminated from high school. *American Journal of Public Health* 70 (6):599-603.

- Robertson, L. S., and P. L. Zador. 1978. Driver education and fatal crash involvement of teenage drivers. *American Journal of Public Health* 68 (10):959-965.
- Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2004. *Evaluation. A systematic approach*. 7th ed. Thousand Oaks, CA: Sage Publications.
- RSEA (Road Safety Educators' Association). 2005. *Process for curriculum review for Driving School Association of the Americas*. Ontario, Canada: RSEA.
- Scriven, M., quoted in Davidson, E. J. 2004. *Evaluation methodology basics*. Thousand Oaks, CA: Sage Publications.
- Shope, J. T., and L. J. Molnar. 2003. Graduated driver licensing in the United States: Evaluation results from the early programs. *Journal of Safety Research*, 34 (1S):63-69.
- Siegrist, S. 2003. *Driver training and licensing—The European perspective*. Paper presented at the NTSB Forum on Driver Education and Training, October, 2003. Washington, DC: National Transportation Safety Board.
- Simpson, H., ed. 1996. *New to the road: Reducing the risks for young motorists*. First Annual International Symposium of the Youth Enhancement Service, June 8-11, 1995: Youth Enhancement Service.
- Sleet, D.A. & Lonero, L. (2002). Behavioral approaches to preventing motor vehicle crashes. In L. Breslow, J. Last, L. Green, & M. McGinnis (Eds.), *Encyclopedia of Public Health*. Mosby.
- Smiley, A., L. Lonero, and M. Chipman. 2004. *Final report. A review of the effectiveness of driver training and licensing programs in reducing road crashes*. Paris, France: MAIF Foundation.
- Smith, D. L. 1983. The DeKalb driver education project—The same mistakes: Improper criteria. *Journal of Traffic Safety Education* 30 (January):14 & 26.
- Smith, M. F., and J. Blatt. 1987. *Summary of preliminary results: Follow-up evaluation. Safe Performance Curriculum Driver Education Project*. Spokane, WA: American Driver & Traffic Safety Education Association.
- Stock, J. R., J. K. Weaver, H. W. Ray, J. R. Brink, and M. G. Sadof. 1983. *Evaluation of Safe Performance Secondary School Driver Education Curriculum Demonstration Project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

- Strang, P. M., K. B. Deutsch, R. S. James, and S. M. Manders. 1982. *A comparison of on-road and off-road driver training*. Canberra City, Australia: Department of Transport and Construction Office of Road Safety.
- Stufflebeam, D. L. 2001. Evaluation models. *New Directions for Evaluation* (89):1-106.
- Stufflebeam, D. L. 2003. The *CIPP model for evaluation*. Paper presented at Annual Conference of the Oregon Program Evaluators' Network, October 3, at Portland, Oregon.
- THCU (The Health Communication Unit). 2002. *Evaluating health promotion programs*. Toronto, ON: Centre for Health Promotion, University of Toronto.
- Trochim, W. 2001. *The research methods knowledge base*. Cincinnati, OH: Atomic Dog Publishing.
- Vernick, J. S., G. Li, S. Ogaitis, E. J. MacKenzie, S. P. Baker, and A. C. Gielen. 1999. Effects of high school driver education on motor vehicle crashes, violations, and licensure. *American Journal of Preventive Medicine* 16 (1S):40-46.
- Weaver, J. K. 1987. *Summary of preliminary results. Follow-up evaluation. Safe Performance Curriculum Driver Education Project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Wiggins, S. 2005. *Graduated Licensing Program interim evaluation report - Year 3*. Vancouver, BC: Insurance Corporation of British Columbia.
- Wilde, G. J. S. 1993. Improve risk perception and reduce risk acceptance: Two proposals for driver education. In *Novice Driver Education Symposium, University of Alberta, Edmonton, Alberta, April 22-33*. Edmonton, AB: Alberta Motor Association Driver Education.
- Wilde, G. J. S. 1994. Risk homeostasis theory and its promise for improved safety. In *Challenges to accident prevention: The issue of risk compensation behaviour*, edited by R. M. Trimpop and G. J. S. Wilde. Groningen, The Netherlands: Styx Publications.
- Williams, A. F., & O'Neill, B. (1974). On-the-road driving records of licensed race drivers. *Accident Analysis & Prevention*, 6, 263-270.
- Williams, A. F. 2003. Teenage drivers: Patterns of risk. *Journal of Safety Research*, 34 (1S): 5-15.
- Williams, A.F., Preusser, D.F. and Ledingham, K.A. (2009). *Feasibility Study on Evaluating Driver Education Curriculum* U.S. Department of Transportation National Highway Traffic Safety Administration DOT HS 811 108.

- W. K. Kellogg Foundation. 2004. *W.K. Kellogg Foundation evaluation handbook*. 1998 [accessed November, 2004]. Available from <http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf>
- Woolley, J. 2000. *In-car driver training at high schools: A literature review*. Adelaide, Australia: Transport SA.
- Wynne-Jones, J. D., and P. M. Hurst. 1984. The AA driver training evaluation. In *Wellington, New Zealand: Ministry of Transport*. Traffic Research Report No. 33. Wellington, New Zealand: Ministry of Transport.
- Zhao, J., R. E. Mann, M. Chipman, E. Adlaf, G. Stoduto, and R. G. Smart. 2006. The impact of driver education on self-reported collisions among young drivers with a graduated license. *Accident Analysis and Prevention* 38:35-42.