CAC Exhibit _____

★ Teen Driver Safety

Large-Scale Evaluation of Driver Education
Review of the Literature on Driver Education Evaluation
2010 Update

Lawrence Lonero
Northport Associates

Dan Mayhew
Traffic Injury Research Foundation

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

NHTSA
www.nhtsa.gov

Manitoba Public Insurance

AAA Foundation for Traffic Safety

607 14th Street, NW, Suite 201 | Washington, DC 20005 | AAAFoundation.org | 202-638-5944

Contents

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

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 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).

[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.

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.

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 (Lonero 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

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);

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.

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 (÷=.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.

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.

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

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.

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 multivariate 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

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

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

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 *California*	• 2,057 students randomly assigned to two training conditions	• Those receiving range practice had fewer recorded crashes, but tests scores were no different	• Randomized control trial • Intermediate measures • No follow-up survey for exposure & behavioral measures
Ray et al. 1980 Stock et al. 1983 *DeKalb County, Georgia*	• Intensive, minimal, and no driver education groups • About 6,000 students randomly assigned to each group	• Intensive training (SPC) drivers had better skills and fewer crashes during first 6 months, but not beyond • Effects were complex; see text	• Comprehensive randomized controlled trial • Long follow-up – 6 years • Formative evaluations and intermediate outcomes measures
Wynne-Jones and Hurst 1984 *New Zealand*	• 788 students, 561 received course, 227 family/friend taught • Random assignment	• No reduction in collisions for driver education group	• Adequate design • Small control group • No formative evaluation or intermediate outcomes
Gregersen 1994 *Sweden*	• 850 students received driver education course compared to controls • Random assignment	• Driver education group significantly worse first year, significantly better second year	• Longer follow-up – 2 years • Reasonable sample size
Masten and Chapman 2003; 2004 *California*	• 1,300 students randomly assigned to one of four instructional settings	• Home-based methods better for 1 knowledge and attitude test, classroom better for DMV knowledge test	• Sample size adequate • Well planned and controlled • Psychometric measures only
Quasi-Experimental Follow-Up Studies			
Forsyth et al. 1995 *United Kingdom*	• Survey of 15,000 new drivers	• Longer time learning to drive associated with fewer crashes for	• Several follow-ups over time • Self-selection bias

Reference	Design	Results	Methodological Strengths/Limitations
47 States	driver education	fatal crashes	
Carstensen 2002 *Denmark*	• Mandatory driver education, new curriculum	• Reduced crashes	• Large sample size • No control of confounding variables

6. 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

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.

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.

7. 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.

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.

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. Schmatz. 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. DiGuiseppi, 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).

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) Geller, 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.

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.

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.