

COMPETENCIES AND INTEREST IN A PROBLEM-FOCUSED UNDERGRADUATE RESEARCH METHODS CRIMINAL JUSTICE COURSE: TWO ASSESSMENTS*

**RALPH B. TAYLOR
TODD ANDERSON
PATRICK MCCONNELL**
Temple University

This study reports on two efforts, in a problem-centered undergraduate criminal justice research methods course, to gauge students' skills in understanding graphical data display and quasi-experimental designs. The problems addressed were drunk driving and handguns. Results show a moderate level of skill by students at course conclusion, but the weak quasi-experimental design precludes confirming the skills were acquired because of the course itself. Students did, however, report increased interest in the two problems because of course coverage. The approach described here may help bridge the gap between skill vs. content vs. process centered views on criminal justice as a liberal arts major.

If criminal justice is a liberal arts undergraduate major (Flanagan 2000), then our reflections on what we teach in our courses should be informed by broader discussions about the goals of liberal arts undergraduate instruction. Over the past several decades three divergent perspectives have emerged on what these goals should be: content, competencies, and process. We will summarize each of these below. Currently the competencies or skill-based position is one of the most popular here in the United States and in Britain. Building on this outlook, this paper reports the results of two independent, end-of-semester assessments, a year apart, of competencies and interest in a problem-focused, undergraduate criminal justice research methods course. The course introduced students to two specific, policy-relevant, criminal justice issues, and used these issues, reflected in

* The views expressed here are solely those of the authors and do not necessarily reflect the views of Temple University, the College of Liberal Arts, or the Department of Criminal Justice. Ruth Karras, Associate Dean, College of Liberal Arts, provided support to the Department of Criminal Justice during spring semester 2000 for a competencies assessment focused on Introduction to Criminal Justice. That support, and the resulting activities, contributed substantively to this effort. Associate Dean Sherri Grasmuck's subsequent support and encouragement also helped fuel the current process. The editor, three helpful reviewers, Pat Jenkins and Ron Davis provided extremely useful comments on earlier drafts. Request reprints from RBT, Department of Criminal Justice, Gladfelter Hall, Temple University, 1115 West Berks Street, Philadelphia, PA 19122 (email: breck@rbytaylor.net).

the literature, class discussions, national data, and in students' responses to a mass questionnaire, as a vehicle for introducing research methods topics. We will learn if students, upon completing this course, can demonstrate specific research methods skills, and if their inquisitiveness about two difficult criminal justice-related problems has increased.

Three Perspectives

The idea that liberal arts was an end in itself (Mann 1999), with the end usually gained by mastering specified content in specified areas (Flanagan 2000), dates from the 18th and 19th centuries. The idea of an intrinsically valuable liberal arts education — learning for learning's sake, appropriate for a large segment of society — has ceded ground, however, to mounting numbers of practically-minded critics for the past three if not five decades. The massive increase in higher education opportunities arising from the post-World War II GI bill (Bennett 1994), and the accompanying increased emphasis on higher education as a bridge to work (Kerr 1994), fueled these criticisms. In the United Kingdom the rationale for liberal arts courses shifted in the 1970s to a focus on competencies, also called core or key skills (Drew 1998; Mann 1999). The skills or competencies rationale provided a second perspective on the value of liberal arts curricula. "This was the start of the decline of education being viewed as an end in itself, available only to the wealthy who could afford such educational advancement without it necessarily leading to career advancement. This was the start of the replacement of 'liberal education' for the few with 'skills approaches' for the many" (Mann 1999:439).

On the policy side, a focus on competencies and their assessment guide and inform curriculum reform and review. These activities build on stating the competencies specific to each course, as well as to the entire curriculum, gauging progress towards acquiring those competencies, revising courses and curricula to remedy deficiencies, and re-assessing (Diamond 1998). Historically, the competencies approach represents an enlargement of the polytechnic model, first emerging in the 1830s to 1850s alongside the new technical schools in the U.S. and Europe; the model sought to provide skilled labor for the industrial revolution (Katz 1975).

Increasing attention to competencies has raised assessment questions: How do we know if students have acquired competencies? How can we be sure it was due to the course or courses and not something else? How do we decide what competencies are most important?

We cannot use course evaluations or grades. Although end of semester evaluations do suggest dimensions of teacher effectiveness (Patrick and Smart 1998), they reflect far more than course content, instructor personality, and presentation effectiveness. Teacher ratings are influenced by received and expected grades (Hamilton 1980; Greenwald 1997; Greenwald and Gillmore 1997), gender of instructor and students (Dukes and

Victoria 1989; Tieman & Rankin-Ullock 1985), instructor status (Hamilton 1980), interactions between instructor and student characteristics (Dukes & Victoria 1989), and broader factors such as class size (Hamilton 1980). Grades themselves are not all that helpful either because they reflect interactions between students' learning styles and instructors' teaching approaches, as well as course content learned (Hedley 1978). We expand below on this question of competencies assessment approaches.

The third approach focuses on psychological and behavioral transformations spurred by the students' thoughts and concerns emerging in the process of learning. Liberal arts education should generate genuine, deep interest in a problem, issue or idea, and enhance skills employed to further this interest. Such changes may alter career decisions and societal contributions made by individuals over the life course (Freedman 1996). Evidence of the impacts described by this view only will reveal themselves over the adult life course of the student. But, in the short term, enhanced student interest in an issue or problem, and/or motivation to learn more about it, seems a relevant outcome.

Pursuing the competencies perspective, we can ask how much agreement there is on the relevant outcomes. Many educators criss-crossing the country argue that we have agreement on the skills needed by the successful college graduate: "communicating . . . mathematics . . . problem solving, and critical thinking; interpersonal skills . . . computer literacy . . . appreciation of cultural diversity" (Diamond 1997: A14). But instructors in a discipline must translate these general goals into more specific skill sets for the overall major, and then determine which of those is relevant to which course.

Flanagan (2000) identifies four core themes deserving emphasis in criminal justice curricula; if these are emphasized, he suggests, a more successful integration between criminal justice and other liberal arts disciplines would result. One of those themes touches on "a frame for analyzing the change process – at the individual, organizational, institutional, and community levels" (Flanagan 2000:8). Part of developing that framework is helping students develop "a keen appreciation of the properties and limits of the scientific method" (Flanagan 2000:9).¹

In our undergraduate major at Temple, three courses address this theme: planned change, introduction to statistics, and introduction to research methods. In two different sections of the last course, held a year apart, we asked students to answer questions intended to demonstrate their

¹ Flanagan's (2000) three other themes are: a focus on "social control in organized society" (p. 7); exploring "moral questions involving fundamental concepts such as integrity, justice, fairness, and equity" (p. 8); and examining "how societies construct organizations and institutions to achieve desired social goals" (p. 8). Given the focus in this paper on an undergraduate research methods course, these three other themes are not relevant. Flanagan's description, however, does raise the interesting question of how we would gauge students' increased competencies as a result of courses that intentionally wrestle with these broader issues.

understanding of this theme. The two specific topics addressed were interpretations of graphical data display (Taylor 1994: Chapter 5) and ability to understand the research methods section of a journal article describing a quasi-experiment (Taylor 1994: Chapter 13).

The implementation of the research methods course relied upon two criminal justice-related problems of considerable immediacy to undergraduates: drunk driving and handguns. The specific problems provide a common set of issues around which to structure discussions of major ideas in research methods: concepts, variables, operationalization, predictors, outcomes, propositions, hypotheses, measurement error, surveys, experiments and so on. In keeping with the process perspective, we also wanted to see if students' interest in these problems increased as a result of the course. By using *two* specific criminal justice problems we apply a spiral approach, widely documented in secondary school science education for at least three decades (Downing 1995; Murphy 1973). Basically, students get two passes at applying core research methods concepts. Using problems as a vehicle for understanding research methods reverses the suggestion to inject methods and statistics into content courses (Gulley 1982; Johnson & Steward 1997; Markham 1991).

In sum, in keeping with the ideas that the criminal justice undergraduate major is at heart a liberal arts major and that students should be acquiring specific skills in these courses, we will seek to document students' acquisition of specific relevant competencies at the conclusion of a research methods course using two criminal justice-related problems as a vehicle for introducing research methods topics. We also will gauge students' interest in these problems, in keeping with the process perspective on liberal arts education. We will complete the assessment twice, using identical outcome variables, in two offerings of the same course a year apart.

METHODS

Design

In each course, the competencies assessment was administered during the final exam period so students would have no conflicts in attending. Students were told that completing the assessment earned them ten percent of their final course grade, and performance on the assessment would *not* be graded. It took students between 40 and 80 minutes to complete the assessment. This is a replicated, post-test only design (Cook and Campbell 1979) and subject to several potential threats to internal validity. We discuss those limitations, and the limits of alternative designs, in the discussion. The overall course design was similar in the two offerings, although there were differences in readings, exercises, and coverage.²

²Contact the first author for more information on these differences.

Participants

Spring 1999. 58 students completed the post test. 32 (56%) were female 25 (43%) were male; gender was missing for one respondent. Ages ranged from 19 to 44; almost half the respondents (49%) were 21 or younger; 9% of the respondents were 30 or older. Of the 50 students providing ethnic information, 15 (30%) were African-American, two were Hispanic (4%), four were Asian-American (8%), and 29 (58%) were white.

Spring 2000. 72 students completed the post test. Of the 64 reporting gender, 34 (53%) were male and 30 (47%) were female. Ages ranged from 18 to 51 and were reported by 59 students; slightly over half those reporting (58%) were 21 or younger; 5% were over 30. Of the 64 students supplying ethnic information, 22 (34%) were African-American, two (3%) were Hispanic, four (6%) were Asian-American, 29 (45%) were white, and seven (11%) were other or multi-ethnic.

Dependent Variables

Comprehension of Quasi-Experimental Design. Several questions gauged students' comprehension of a specific quasi-experiment. Students read the entire research design section from an article describing a national evaluation of a delinquency prevention program (GREAT: gang reduction education and treatment) (Esbensen & Osgood 1999). The evaluation design in the study was "post-test only comparison between students who participated in the GREAT program the previous year and students who did not" (Esbensen & Osgood 1999:202). Also included was the section of the paper on the outcome analysis, describing the various factors researchers planned to control so as to increase experimental and control comparability. Seven questions about the study followed, all in true-false format:

1. In this evaluation of GREAT, students were randomly assigned to either participate in the program or to not participate in the program (*Real experiment*).
2. In this evaluation of GREAT, students who had participated in GREAT were compared to a group of non-participating students who were almost exactly equivalent to the participating group (*Equivalence*).
3. The schools where the students answered the questionnaires were a representative random sample of public schools ; thus, the results obtained here can be justifiably generalized to the population of students in public schools in the continental United States (*Representativeness*).
4. In this study the researchers "controlled for" several background factors. They did this to try and reduce pre-existing differences between the treatment and control groups (*Controls*).
5. In the study, if the researchers had found significant differences between treatment and control students on the outcomes, and they had NOT controlled for background

differences between the two groups, these background differences would represent threats to internal validity (TIV) of the results; the background differences, rather than the program, might have been responsible for the differences on the outcomes between experimentals and controls (TIV).

6. The study relies on one and only one outcome variable (*Outcome*).
7. One of the major outcomes examined in this study is whether or not the respondent reports being a gang member (*Membership*).

Comprehension of a Bar Chart. Students were presented with a clustered bar chart, based on 1972-1996 national data from the General Social Survey (GSS), a national survey completed yearly. In both classes, as part of our review of work on guns, students had developed hypotheses about handgun ownership, and had tested their hypotheses with these GSS data. We had reviewed the resulting bar graphs over a couple of class periods. Therefore, students were familiar with the dependent variable, the independent variables, and the data source.³ The results were portrayed in the form of simple bar charts. I encouraged students to use a rough rule of thumb when interpreting the bar chart results: if two percentages differed by more than 5%, they should think of that difference as "significant." When students' hypotheses about the predictors of handgun ownership failed to gain support from the GSS data, I encouraged them to think more carefully about the topic and suggest an additional variable that might condition the expected relationship; i.e., what other factors might be involved? After they sent me their modified hypotheses, I generated clustered bar charts, showing the relationships of both independent variables to the outcome, the percentage owning a handgun. In class we had reviewed a couple of these two-way or clustered bar charts. The particular bar graph shown in the post-test, however, had neither been viewed in class nor shown in class.⁴

In both classes the amount of time devoted to hypothesizing about the determinants of gun ownership, the overall structure of the exercises, and approach to interpreting graphically displayed results, were comparable.

The figure students were asked to interpret in the competencies assessment was introduced as follows:

Take a look at Figure 1 that appears at the end of the assessment. It is based on data from the General Social Survey, 1972 through 1996 (added together). This is a nationally representative survey done yearly. The respondents represent a cross section of households in the United States. The dependent variable is pistol ownership — whether the person reports owning a pistol or handgun. The height of each bar shows you the proportion in that group

³An outline of the basic exercise can be found online at: <http://blue.temple.edu/~ralph/gunex.html>. That page also shows the hypotheses that the students developed in their work groups, and provides links to click on for showing the results.

⁴It can be found on-line at <http://www.rbtaylor.net/gunexrace2.htm>.

reporting that they own a pistol. Respondents are classified in two ways. First, they are separated into two racial groups: African-American vs. not African-American. (In the chart “not African-American” is labeled WHITE.). Second, they are separated into those that live in big cities vs. those that do not. “Pistol” refers to any type of handgun.

The students were asked to answer the following true-false questions based on their examination of the chart:

1. A higher proportion of African-Americans report pistol ownership in the big cities than outside the big cities.
2. A higher proportion of whites report pistol ownership in the big cities than outside the big cities.
3. Among those living outside the big cities, a slightly higher proportion of whites than African-Americans report pistol ownership.
4. Among those living in the big cities, a slightly higher proportion of whites than African-Americans report pistol ownership.
5. The effects of race on pistol ownership are reversed when we switch from those living outside the big cities to those living in the big cities.
6. Your colleagues who were hypothesizing about these data suggested that African-Americans were more likely to own handguns in the big city than whites, because they lived in more dangerous neighborhoods than did the whites. It looks as if these data support that hypothesis.

Interest. The competencies assessment included several items to gauge if students interest in the problems introduced in the course had increased as a function of that exposure (e.g., “Because of this class I am more interested in the problem of drunk driving than I was at the beginning of the semester.”) Items also asked if, as a result of exposure, they had a “better appreciation” of the problem “from a social science perspective.” Finally, questions asked if students felt the coverage of these problems had crowded out other topics, or if they felt the instructor’s presentation of these issues was biased. Response options included: (1) disagree strongly / (2) disagree / (3) disagree slightly / (4) agree slightly / (5) agree / (6) agree strongly. Some items were arranged so that “agree” was not always the socially desirable response.

RESULTS

Interpreting Quasi-Experimental Research Design

All these questions were true/false items. The percent correct was calculated with missing data included in the total. See Table 1.

Spring 1999. Across the seven true-false items, respondents averaged 61% correct (median=66%); the percent correct per item ranged from a

Table 1. Comprehension of Quasi-Experiment Methods Section.

Course	Spring 1999	Spring 2000
Question	Percent (n/N) Correct (answer)	
Real experiment	63.8 (37/58) (false)	77.8 (56/72) (false)
Equivalence	43.1 (25/58) (false)	40.3 (29/72) (false)
Representativeness	37.9 (22/58) (false)	39.4 (28/71) (false)
Controls	65.5 (38/58) (true)	81.9 (59/72) (true)
TIV	72.4 (42/58) (true)	79.2 (57/72) (true)
Outcome	74.1 (43/58) (false)	73.2 (52/71) (false)
Membership	70.7 (41/58) (true)	80.6 (58/72) (true)
<i>Across all Seven Items</i>		
Average Percent Correct:	61%	77.8
Median Percent Correct	66%	67.5
Comparison Against Chance	$z = 1.68; p < .05$	$Z = 3.17; p < .001$

Note: For wording of specific items, see text.

low of 38% to a high of 74%.⁵ Students seemed to score best on items concerned with the outcome variable (*outcome, membership*) and threats to internal validity (*TIV*), with over 70% getting all of these correct. Students fared most poorly, about 40% correct, with two items asking about the broader design: Was the control group comparable (*equivalence*)?; Was the sample a representative, random sample of public schools (*representativeness*)?

Do these results represent "success" or not? The z test for a difference in proportions (Blalock 1979: 195-199) suggested that the students' average of 61% correct, across the seven items, significantly exceeded chance guessing ($z = 1.675, p < .05$). Clearly, however, there is still room for considerable improvement.

Spring 2000. The percentage answering correctly for the various items ranged from a low of 39% to a high of 82%; the proportion getting the correct answer, averaged across all the items was 68%. A z test for a difference in proportions showed students did far better than just random guessing ($p < .001$).

Even though the spring 2000 overall average correct was slightly higher, the pattern of best handled vs. worst handled items generally paralleled the spring 1999 pattern. The two questions asking about inferences from the design (*Equivalence* and *Representativeness*), the most poorly handled by the spring 1999 students, also had the lowest percentage of correct answers in spring 2000 (40% and 39%, respectively). The three items most often correctly answered by the spring 1999 students were

⁵These percentages were calculated with missing data included in the totals and thus treat no answer as a wrong answer.

among the five most often correctly answered by the spring 2000 students, and addressed the dependent variable (*Outcome, Membership*) and internal validity (*TIV*).

Differences between the two groups surfaced for two items. Spring 2000 students did better than spring 1999 students on understanding the quasi-experimental nature of the design (*Real experiment*, 78% correct in spring 2000 vs. 64% in spring 1999), and the question about covarying out confounds (*Controls*; 82% in spring 2000 vs. 65% in spring 1999).

In two different class offerings, students can correctly answer several items about a quasi-experimental design, at a rate better than chance guessing, based on their reading of a previously unread journal article methods section. Further, the best and worst handled items overlapped substantially in the two groups of students.

Bar Chart Comprehension

Spring 1999. The questions and the percent correct responses appear in Table 2. Across the six items the proportion supplying the correct answer ranged from 81% to 62% and averaged 70.7% (median=70.7% also). Results from the z test of the difference in proportions suggest students did significantly better than random guessing on these items ($z=4.11$; $p < .001$).

Three of the six statements were correctly classified by at least 70% of the respondents: higher proportion of handgun-owning whites in big cities than outside (72%); outside big cities higher proportion of whites as compared to African-American handgun owners (74%); and effects of race on ownership depend on big city vs. elsewhere (81%).

Spring 2000. The results on bar chart comprehension were virtually identical to those from a year earlier. Averaged across all the items, about 70% of the students provided a correct answer, significantly better than chance guessing ($z=4.37$; $p < .001$). Again, there were parallel points in the pattern across items. Two of the three items correctly answered by 70% or more of the spring 1999 students were also correctly answered by 70% or more of the spring 2000 students. The item most frequently classified correctly was the same in both courses: effects of race on pistol ownership depend on city size (81% correct in spring 1999; 93% correct in spring 2000).

Evaluative Reactions and Problem Interest

Students in both classes slightly agreed that at the semester's end they had a better appreciation of the problem of drunk driving (means = 4.44 and 4.76 for 1999 and 2000), between "agree slightly" (4) and "agree" (5). For the problem of guns the average answers were roughly comparable; the means were 4.25 for 1999 and 4.49 for 2000. So it appears that in both courses students on average reported that their insight into these pressing problems had been somewhat enhanced. See Table 3.

Table 2. Comprehension of Clustered Bar Chart.

Course	Spring 1999	Spring 2000
Item	% correct (n/N) (answer)	
A higher proportion of African-Americans report pistol ownership in the big cities than outside the big cities.	69% (40/58) (false)	69.4% (50/72) (false)
A higher proportion of whites report pistol ownership in the big cities than outside the big cities.	72.4% (42/58) (false)	70.8% (51/72) (false)
Among those living outside the big cities, a slightly higher proportion of whites than African-Americans report pistol ownership.	74.1% (43/58) (true)	63.9% (46/72) (true)
Among those living in the big cities, a slightly higher proportion of whites than African-Americans report pistol ownership.	62.1% (36/58) (false)	68.1% (49/72) (false)
The effects of race on pistol ownership are reversed when we switch from those living outside the big cities to those living in the big cities.	81% (47/58) (true)	93% (66/71) (true)
Your colleagues who were hypothesizing about these data suggested that African-Americans were more likely to own handguns in the big city than whites, because they lived in more dangerous neighborhoods than did the whites. It looks as if these data support that hypothesis.	65.5% (38/58) (true)	72.2% (52/72) (true)
<i>Across all six items</i>		
Average Percent Correct:	70.7%	70.9%
Median Percent Correct:	70.7%	70.1%
Comparison Against Chance: z=	4.11; p < .001	4.37; p < .001

Students also reported slight increases in their interest in the problem areas themselves. When asked if they were "more interested" in either guns or drunk driving than they were at the beginning of the semester, *because of the course*, average responses for the gun issue were above "agree slightly" for both classes (mean = 4.19 for 1999, 4.43 for 2000). For the drunk driving problem, 1999 students' average response fell just below "slightly agree" (mean=3.86) and 2000 students' average response fell just above "slightly agree" (mean=4.14). But for three out of these four interest

Table 3. Evaluative Reactions, and Assessments of Interest.

Item	Spring 1999	Spring 2000
	mean (se) n	mean (se) n
In this course we spent too much time learning about drunk driving.	2.86 (.15) 58	2.73 (.14) 70
In this course we spent too much time learning about guns.	2.56 (.16) 57	2.46 (.12) 70
I think this course would have been a lot more interesting if the professor had just concentrated on research methods, and NOT spent so much time teaching us about guns and drunk driving.	2.24 (.19) 55	2.20 (.16) 66
As a result of this course, I have a better appreciation of the problem of drunk driving from a social science perspective.	4.44 (.14) 57	4.76 (.12) 71
As a result of this course, I have a better appreciation of the costs and benefits of guns from a social science perspective.	4.25 (.14) 57	4.49 (.14) 72
The professor was UNable to present information about drunk driving effectively because his biases on the topic were so strong.	1.96 (.13) 55	1.72 (.11) 69
The professor was UNable to present information about guns and handguns effectively because his biases on the topic were so strong.	1.89 (.13) 54	1.71 (.10) 69
Because of this class I am more interested in the problem of drunk driving than I was at the beginning of the semester.	3.86 (.16) 58	4.14 (.15) 72
Because of this class I am more interested in the issue of guns and handguns than I was at the beginning of the semester.	4.19 (.14) 42	4.43 (.14) 68

Note: The response format used was: 1 = disagree strongly; 2 = disagree; 3 = disagree slightly; 4 = agree slightly; 5 = agree; 6 = agree strongly.

items, the means were at least two standard errors above a response indicating no change in interest.⁶

Another way to get at students' reactions to the two problems used was to ask them about the amount of time spent on the topic. For both years and for both topics, students *disagreed* that "in the course, we spent too much time learning about" drunk driving or guns. The drunk driving

⁶A score of 3.5, midway between "slightly disagree" (3) and "slightly agree" (4) would reflect no change in interest since the beginning of the semester.

means were 2.86 (1999) and 2.73 (2000), between "disagree" (2) and "disagree slightly" (3). The means for guns were 2.56 in 1999 and 2.46 in 2000, about halfway between "disagree" and "disagree slightly." So students in neither semester viewed the amount of time devoted to these topics as excessive.

A final way to get at students' thoughts about the social problems and related content was to ask students if they thought "the course would have been a lot more interesting if the professor has just concentrated on research methods and NOT spent so much time teaching us about guns and drunk driving." For both semesters the mean response was around "disagree" (2) (1999: 2.24; 2000: 2.20). So students on average did *not* feel that the problem content crowded out fundamental research method issues.

In sum, it appears that students in both semesters expressed increased appreciation of the complexities of these two problems, expressed slightly increased interest compared to the beginning of the semester, and did not feel that the coverage of the problems "shortchanged" them on core topics in research methods.

DISCUSSION

The present results suggest that criminal justice majors in a required undergraduate methods course at a large, diverse, state-aided university have acquired some competence by the end of the semester in interpreting and understanding the strengths and weaknesses of a quasi-experimental research design and in interpreting a moderately complex graphical display of results from a survey. Results also show self-reported increases in interest in the problems addressed. Results replicate across course offerings. Before discussing the implications of these results, we review of the strengths and weaknesses of the current study.

Each study represents a one group, post test-only design (Cook and Campbell 1979: 96-98). This design is potentially subject to numerous plausible threats to internal validity (Taylor 1994: 266-291). Alternate designs, however, were also problematic or not feasible. A one-group pretest-post test design (Cook and Campbell 1979: 99-103) would allow us to see how much individual students gained over the semester. But it too is subject to numerous potential threats to internal validity. For example, it would not rule out maturation arising from aging or material learned in other courses, or instrumentation effects, or regression artifacts. Even if the instrumentation effects could be removed by using different examples in the pretest and post test, presuming one could come up with test material of precisely the same difficulty, maturation and regression threats remain. If the instrumentation effects were removed by having only a random half of the students complete the pretest, then the problem of unacceptably low statistical power is introduced; in these classes about 35-40 students regularly attend out of 60-80 students registered. Further, a pretest-post test

design requires using individual identifiers to enable linking the two sets of scores. This seems likely to raise concerns among students about potential loss of confidentiality in an instrument that also asks about sensitive topics like handgun ownership and experiences of driving while intoxicated.

One reviewer suggested a design with a control group. It is not clear this would necessarily be a better approach. If the control group is another offering of the research methods course, the contrast is between two versions of a methods course, not between treatment vs. no treatment. In addition, differences in instructor are introduced as well. If one used a different criminal justice course, the nonequivalency between treatment and control groups becomes problematic. Students often delay taking the methods course until they are close to graduation. In spring 1999 and in spring 2000 61% and 47%, respectively, expected to graduate within a year of completing the course. The portion of seniors or rising seniors is much higher in this course than in other courses, so much so that partialling out this treatment vs. control difference using another undergraduate course becomes unwise. The "likely equivocalities" would be substantial (Cook and Campbell 1979: 199).

One final limitation deserves mention: our outcomes have addressed only some of the content areas covered in an undergraduate research methods course. Future investigations along these lines should expand the set of outcomes to include other content areas such as research ethics, operationalization, and the like.

Standing against the substantial weaknesses of our design are three strengths. Results were replicated, using identical outcome measures, in two successive offerings with different students a year apart. This suggests, but does not confirm, given the design questions, external validity, which is always an empirical question (Taylor 1994: 164-165). For each outcome, multiple indicators were employed. Further, the outcomes were grounded in "real world problems" of substantial interest to criminal justice majors.

Nonetheless, given our design, potential threats to internal validity, especially of history and maturation (Taylor 1994) remain. Given the problems described above, and the comments on alternate designs, the most useful future design may be as follows. Include pretests, administered to a random half of the class. Separate the pretests and post tests from the broader in-class questionnaire asking about personal experiences and opinions, thereby minimizing students' worries about potential loss of confidentiality in the latter questionnaire. Pretests and post tests can expand to include additional research methods topics. Have experts rate the difficulty of varying examples used in pretests vs. post tests, and covary out these differences in the analyses. Circumvent the problem of low statistical power by analyzing different classes jointly using multilevel models (Kreft and de Leeuw 1998; Snijders and Bosker 1999). Data from several assessments can

be joined while simultaneously attending to the grouped nature of the data and individual starting scores.

Turning to the findings themselves, the current approach of grounding a research methods course in specific social problems, as we have done here with guns and drunk driving, simultaneously furthers the goals emerging from both the competencies and process perspectives on liberal arts, and illustrates how to more solidly ground criminal justice courses in a liberal arts framework. The problems increase students' curiosity, one of the liberal arts goals noted by Freedman; with the graphical display exercise, they are more motivated to figure out "what the numbers mean," which is the skill in question.

Students' engagement was further demonstrated in their self-reports of changes in interest. When asked if their interest in the problem had increased *because* of this class, for the gun issue students on average in both semesters at least agreed slightly that it had; for the drunk driving issue students in spring 2000 on average agreed slightly that it had. (Of course, here too, independent pretests and post tests are desirable.) Although we did not investigate it here, past research on course assessments, described above, suggests possible gender and/or race-based differences in interest gains in the problems studied. Future work can explore these differentials. Hopefully that work would build expectations of differential effects on relevant literature specific to the problem. (In the case of guns, for example, see Ludwig and Cook 1998.)

Introducing specific social problems as vehicles through which to gain expertise in research methods topics also allows the instructor to incorporate perspectives from other disciplines studying those same problems. For example, in gauging effects of alcohol on behavior, we introduce public health readings reviewing gender differences in effects of alcohol on blood alcohol content, which leads to gender-specific definitions of binge drinking (Wechsler, Dowdall, Davenport et al. 1995). We demonstrate how this works through an in-class exercise (Taylor and McConnell 2001). Such a strategy follows Flanagan's (2000: 9) suggestion we "consciously incorporate perspectives from other disciplines."

The second set of outcomes where students interpreted the quasi-experimental design of an evaluation of the GREAT program addresses Flanagan's (2000: 8) theme of providing a "frame for analyzing the change process." Quasi-experiments to evaluate programs are ubiquitous in criminal justice: "Scared Straight" (Finckenauer 1982), "DARE" (Rosenbaum 1994), and Boston's "Operation Ceasefire" (Kennedy 1997; Kennedy 1996) are just a few widely known recent examples. Understanding the "properties and the limitations" (Flanagan 2000:9) of these evaluations is pivotal to gauging the worth and limits of programs seeking to reduce willingness to commit prison-worthy felonies, drug abuse, or gun violence. Developing stronger insight into the inherent weaknesses and strengths of these studies

is not only a pivotal part of appreciating the applied social science enterprise, it also begins to inform students about the causes for tensions between science and public policy. Policy makers are often looking for more definite answers than the studies can produce.

In closing, the tension between an interest-enhancement view of undergraduate liberal arts education, as articulated by Freedman and others, and a competencies-acquisition focus, as articulated by Diamond and others, deserves recognition. At the same time there are opportunities within criminal justice undergraduate courses for *blending* these views. Criminal justice is an undergraduate major which can, Flanagan (2000) has argued, incorporate significant liberal arts themes while simultaneously strengthening students' skills as interpreters of social science. This article has documented efforts to promote both views within a research methods course. The relevant liberal arts theme is assessing change and understanding the limits of social science (Flanagan 2000). It does appear that for the skills investigated, students do have some better than guessing competency, and they report their interest in the problems examined, guns and drunk driving, having increased because of the course. A weak quasi-experimental design was employed here because of both classroom limitations (e.g., n of students), and potential interpretive problems that would have persisted even if more rigorous designs were employed. Although the results here replicated across two different classes, a year apart, caution in interpreting results is warranted. The approach here, hopefully employing a stronger design as described above, could be expanded to encompass different course-related outcomes, or it could also be expanded to other undergraduate, "tool-focused" criminal justice classes, such as statistics.

REFERENCES

- Bennett, M. J. 1994 "The Law That Worked." *Educational Record* fall:6-14.
- Diamond, R. M. 1997. "Curriculum Reform Needed If Students Are to Master Core Skills." *The Chronicle of Higher Education* :A14.
- Diamond, R. M. 1998. *Designing and assessing courses and curricula: A Practical Guide*, Second Edition. San Francisco: Jossey-Bass.
- Downing, C. 1995. "Science in a S.a.c.k: A Spiral Approach to Content Knowledge." *Science Teacher* 62:46-48.
- Drew, S. 1998. *Key Skills in Higher Education: Background and Rationale*. London: SEDA.
- Dukes, R. L. and G. Victoria. 1989. "The Effects of Gender, Status and Effective Teaching of the Evaluation of College Instruction." *Teaching Sociology* 17:447-457.
- Esbensen, F. A. and D. W. Osgood. 1999. "Gang Resistance Education and Training (Great): Results from the National Evaluation." *Journal of Research in Crime and Delinquency* 36:194-225.
- Finckenauer, J. O. 1982. *Scared Straight! and the Panacea Phenomenon*. Englewood Cliffs, NJ: Prentice Hall.
- Flanagan, T. J. 2000. "Liberal Education and the Criminal Justice Major." *Journal of Criminal Justice Education* 11:1-14.
- Freedman, J. O. 1996. *Idealism and liberal education*. Ann Arbor: University of Michigan Press.

- Greenwald, A. G. 1997. "Validity Concerns and Usefulness of Student Ratings of Instruction." *American Psychologist* 52: 1182-1186.
- Greenwald, A. G., Gillmore, G. M. 1997. "How Useful Are Student Ratings? Reactions to Comments on the Current Issues Section." *American Psychologist* 11: 1228-1229.
- Gulley, W. 1982. "Integrating Theory, Methods, and Statistics." *Teaching Sociology* 10:65-70.
- Hamilton, L. L. 1980. "Grades, Class Size and Faculty Status Predict Teaching Evaluations." *Teaching Sociology* 8:47-62.
- Hedley, R. A. 1978. "Measurement: Social Research Strategies and Their Relevance to Grading." *Teaching Sociology* 6:21-29.
- Johnson, M. A. and G. Steward. 1997. "Integrating Research Methods into Substantive Courses: a Course Project to Identify Social Backgrounds of Political Elites." *Teaching Sociology* 25:168-175.
- Katz, M. 1975. *Class, bureaucracy, and Schools: The Illusion of Educational Change in America*. New York: Norton.
- Kennedy, D. M. 1996. "Violence in Boston: Gun Markets, Serious Youth Offenders, and a Use-reduction Strategy." *Law and Contemporary Problems* 59:147.
- 1997. "Pulling Levers: Chronic Offenders, High Crime Settings, and a Theory of Prevention." *Valparaiso University Law Review* 31:449-484.
- Kerr, C. 1994. "Expanding Access and Changing Missions: the Federal Role in U.S. Higher Education." *Educational Record* fall:27-31.
- Kreft, I. and J. de Leeuw. 1998. *Introducing multilevel modeling*. Thousand Oaks: Sage.
- Ludwig, J. and P. J. Cook. 1998. "The Gender Gap in Reporting Household Gun Ownership." *American Journal of Public Health* 88:1715-1718.
- Mann, S. 1999. "Review of "Key Skills in Higher Education" (1998), by Sue Drew." *Assessment and Evaluation in Higher Education* 24:439-440.
- Markham, W. T. 1991. "Research Methods in the Introductory Course: To Be or Not to Be?" *Teaching Sociology* 19:464-471.
- Murphy, P. D. 1973. "Modules for Consumer Education: A Spiral-process Approach to Curriculum Development." *American Vocational Journal* 48:7,52.
- Patrick, J. and R. M. Smart. 1998. "An Empirical Evaluation of Teacher Effectiveness: the Emergence of Three Critical Factors." *Assessment and Evaluation in Higher Education* 23:165-179.
- Rosenbaum, D. P. 1994. "Cops in the Classroom: a Longitudinal Evaluation of Drug Abuse Resistance Education (Dare)." *Journal of Research in Crime and Delinquency* 31:3-31.
- Snijders, T. A. B. and R. J. Bosker. 1999. *Multilevel analysis: An Introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage.
- Taylor, R. B. 1994. *Research methods in criminal justice*. New York: McGraw Hill.
- Taylor, R. B., and P. McConnell. 2001. "BAC and Beer: Operationalizing Drunk Driving Laws in a Research Methods Course Exercise." *Teaching Sociology* 29: 219-228.
- Tieman, C. R. and B. Rankin-Ullock. 1985. "Student Evaluations of Teachers." *Teaching Sociology* 12:177-191.
- Wechsler, H., G. W. Dowdall, A. Davenport, and E. B. Rimm. 1995. "A Gender-specific Measure of Binge Drinking among College Students." *American Journal of Public Health* 85:982-985.

