A Meta-Analytic Assessment of Delinquency-Related Outcomes of Alternative Education Programs
Stephen M. Cox, William S. Davidson and Timothy S. Bynum
Crime & Delinquency 1995 41: 219
DOI: 10.1177/0011128795041002004

The online version of this article can be found at:
http://cad.sagepub.com/content/41/2/219

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Crime & Delinquency can be found at:

Email Alerts: http://cad.sagepub.com/cgi/alerts
Subscriptions: http://cad.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://cad.sagepub.com/content/41/2/219.refs.html
A Meta-Analytic Assessment of Delinquency-Related Outcomes of Alternative Education Programs

Stephen M. Cox
William S. Davidson
Timothy S. Bynum

Although the alternative education movement continues to grow, uncertainty is present across the literature regarding the effectiveness of these programs. Prior reviews have found that alternative schools improve school performance, attitudes toward school, school attendance, and self-esteem, while decreasing delinquency. However, these reviews have been unable to determine the magnitude of these effects or examine potential correlates of success. The present study used meta-analysis to quantitatively summarize prior empirical research on alternative schools. The meta-analysis findings show that alternative education programs have a small overall effect on school performance, attitudes toward school, and self-esteem but no effect on delinquency. Furthermore, alternative education programs that target a specific population of at-risk delinquents or low school achievers produce larger effects than programs with open admissions.

There is an extensive literature dedicated to developing the relationship between schools and misbehavior (e.g., Sampson and Laub 1993; G. Gottfredson 1987; Lawrence 1985; Polk 1984; Pink 1982; Liazos 1978). Studies have found that a variety of school-related variables are causally linked to delinquent behavior, including: school performance (Jarjoura 1993; Farrington 1992; Tremblay et al. 1992; Cohill 1991; Tracy, Wolfgang, and Figlio 1990; Short 1990; Phillips and Kelly 1979), school attendance (Thornberry, Moore, and Christenson 1985; Elliott and Voss 1974), and attitudes toward

STEPHEN M. COX: Doctoral candidate, School of Criminal Justice, Michigan State University.
WILLIAMS DAVIDSON: Professor, Department of Psychology, Michigan State University.
TIMOTHY S. BYNUM: Professor, School of Criminal Justice, Michigan State University.

The authors would like to thank Don C. Gibbons for his extremely detailed review of an earlier draft. Other reviewers who were very helpful included Jennifer Hedlund, Beverly Dickinson, and Madeline Wordes.

CRIME & DELinquency, Vol. 41 No. 2, April 1995 219-234
© 1995 Sage Publications, Inc.
school (Loeber, Stouthamer-Loeber, Van Kammen, and Farrington 1991; Mak 1991; Sederstrom and Weis 1981; Kelly and Balch 1971). The relationship between schooling and delinquency is strong and provides a foundation for developing strategies for preventing delinquency (Gottfredson 1981). One such strategy has been alternative education.

Although there are no standard models for alternative schools, they are generally designed to create a more positive learning environment through low teacher-to-student ratios, individualized and self-paced instruction, non-competitive performance assessments, and less-structured classrooms (Raywid 1983). The belief is that a self-paced curriculum and an informal classroom structure allow students to work independently and afford staff more time for individualized instruction. Alternative school students are also believed to be under less pressure to perform at the same level as other students because success is measured by individual achievements, rather than by comparing an individual’s progress to that of the entire class (Gold and Mann 1984).

In theory, students feel comfortable in this environment and are more motivated to attend this type of school. Students attending alternative schools are believed to have higher self-esteem, more positive attitudes toward school, improved school attendance, higher academic performance, and decreased delinquent behaviors than when they attended traditional schools. Based on these beliefs, many alternative education programs have targeted delinquent youth (D. Gottfredson 1987; Arnove and Strout 1980).

PRIOR REVIEWS

There are generally two divergent conclusions from literature reviews of alternative school programs:

1. These programs have been effective in achieving positive outcomes in student attitudes, academic achievement, self-esteem, and student behavior (Young 1990; Garrison 1987; Reilly and Reilly 1983; Barr, Colston, and Parrett 1977).
2. It was not possible to reach conclusions regarding the effectiveness of alternative schools or factors that correlate with success (Hawkins and Wall 1980; Duke and Muzio 1978).

Also, earlier reviews have described characteristics of successful programs and have noted common limitations of individual studies.
CHARACTERISTICS OF SUCCESSFUL ALTERNATIVE SCHOOLS

Young (1990) reviewed several empirical studies of alternative schools and concluded that small school size, a supportive and noncompetitive environment, and a student-centered curriculum were structural characteristics commonly associated with program success. Although there is general agreement on the structural characteristics of successful alternative schools, there is little consensus regarding the appropriate population for these programs. For instance, Reilly and Reilly (1983) contend that students already performing at high levels of academic achievement will continue to be successful in an alternative school, whereas lower-achieving students can be expected to be unsuccessful, especially if they have attendance and/or disciplinary problems. In contrast, others have concluded that low school achievers should perform better in alternative schools than in traditional ones. Moreover, programs focusing on specific target populations are more likely to produce positive results than are less focused programs, because the program structure will be centered on a definitive problem or need (Young 1990; Garrison 1987; and G. Gottfredson 1987).

Previous literature reviews have also been unable to arrive at definitive conclusions regarding the impact of racial heterogeneity and school success. Although alternative education programs should reflect the cultural diversity of the community in which they are located (Duke 1978), many alternative schools are dumping grounds or warehouses for social misfits and academically incompetent students. Arno and Strout (1980) have argued that minority students are often overrepresented in these programs, as they have a higher percentage of suspensions and exclusions from school due to the discriminatory practices of many school districts. Such practices place the programs in danger of becoming racially isolated from the community and could inhibit the effectiveness of the school. Studies assessing these claims have been inconclusive (Reilly and Reilly 1983).

Methodological Limitations of Individual Studies

These reviews have also pointed out several methodological shortcomings frequently found in prior research. The problems commonly encountered were:

1. A lack of control or comparison group data
2. Failure to randomize when sampling from populations of students, teachers, and parents
3. A tendency to eliminate data on program dropouts from aggregate statistics or final data analysis
4. A lack of follow-up data on students who leave early or graduate from alternative schools (Duke and Muzio 1978)

These limitations weaken the evaluators’ ability to determine whether the alternative school generated the desired effects (Hawkins and Wall 1980).

Limitations of Prior Literature Reviews

The prior reviews have been limited to a narrative approach, summarizing the alternative education literature. Several of them provided a case-by-case synopsis of individual programs based on prior evaluations or reports, whereas others summarized the findings of published studies categorized by specific outcome variables (e.g., delinquent behavior, academic achievement, attendance, school attitude, self-concept). Caution should be exercised when interpreting traditional literature reviews because they tend to suffer from: (a) selective inclusion of studies, (b) differential subjective weighting of studies in the interpretation of a set of findings, (c) misleading interpretations of study findings, (d) failure to examine other study characteristics as potential explanations for consistent results across studies, and (e) failure to examine the effect of moderator variables in relationship to the outcome variable (Wolf 1986).

THE PRESENT STUDY

Meta-Analysis

In order to overcome the limitations of narrative reviews, the present study employed meta-analysis, a quantitative method of summarizing the findings from empirical research. Meta-analytic procedures include an estimate of the magnitude of the program impact (effect size) and an estimate of errors that may be due to sampling or measurement artifacts (Schmitt and Klimoski 1991; Hunter and Schmidt 1990; Glass, McGaw, and Smith 1981). In addition, meta-analysis allows for coding of intervention characteristics and evaluation methods of the study for the purpose of examining relationships between these characteristics and the quantitative estimate of program impact (Mayer, Gensheimer, Davidson, and Gottschalk 1986). By summarizing findings across studies, this technique can differentiate program effects from other differences, such as uncontrolled characteristics of the sample (assignment of subjects, grade level, target populations, etc.), deficiencies in the research design (e.g.,
whether a pre-post design yields different results than experimental designs), and other sources of variation (e.g., intensity and duration of the intervention) that may not have been adequately controlled in single studies (Logan and Gaes 1993).

**METHOD**

**Sample of Studies**

The present review was based upon investigations obtained from computer searches of the Educational Resources Information Circuit (ERIC), PsychLit, and the National Criminal Justice Reference Service (NCJRS) covering the years 1966 through 1993. These searches yielded 241 citations referring to alternative education programs.

The abstracts of the 241 citations were reviewed, and studies with a program that did not fit the general definition of an alternative school and/or that were not empirical evaluations were omitted. To be included in the database, alternative education programs had to consist of a separate curriculum, be housed outside of the conventional school,¹ and statistically assess at least one type of outcome (e.g., student attitudes toward school, school performance, self-esteem, and delinquency). Only 87 of the studies met these criteria.²

Furthermore, because meta-analysis employs quantitative methods of converting statistical results into a standard metric (Wachter and Straf 1990), studies failing to report or make reference to statistical results (e.g., means, standard deviations, correlations, t tests, chi-squares, F values, etc.) were not included in the present study. Out of the 87 studies, 57 were evaluations of alternative schools that provided adequate statistics necessary for inclusion in the meta-analysis.³

**Coding Methods**

**Study characteristics.** The 57 evaluations were coded according to procedures described by Hunter and Schmidt (1990). Detailed coding was undertaken in order to obtain as much information as possible regarding the study characteristics. The variables included: Sample characteristics (gender ratio in sample, racial composition, age, grade in school, and target characteristics of sample), intervention characteristics (whether the alternative school was in an urban, rural, or suburban area, daily length of program, and overall duration of the program), and methodological characteristics of the study
(type of research design used and how the participants were assigned to the program).

The coded outcome variables were delinquency, school performance, school attitude, and self-esteem. Official delinquency data (e.g., police contacts, juvenile court records, etc.) were combined with self-reported delinquency data for the delinquency outcome variable. School performance was measured by amalgamating standardized test scores, classroom grades, and attendance. Attachment to school, attitude toward school, and commitment to school activities data were combined to create a school attitude variable.

The self-esteem outcome variable was based upon various measures of self-esteem and self-concept. Self-esteem was included in the meta-analysis even though it is generally not considered a school-related variable. There is strong evidence in the delinquency literature that self-esteem is related to delinquency (McCarthy and Hoge 1984; Bynner, O’Malley, and Bachman 1981; Gold 1978; Rosenberg and Rosenberg 1978; Gold and Mann 1972; Reckless, Dinitz, and Murray 1956). In addition, many evaluations of alternative school programs have included self-esteem as an outcome variable.

Assessment of program effects. Two separate methods were used to assess program effects. First, an overall effectiveness rating was assigned to each study based on its overall conclusions. This rating consisted of positive effects, negative effects, and no effects. For example, if the author concluded that the program produced an overall positive effect, the study was assigned an overall rating of positive.

Effectiveness ratings were also assigned to individual program outcomes within each study. These were drawn from the conclusions of the study regarding specific program effects and were coded as having a positive effect, negative effect, or no effect. Using this method, it was possible to assess a program’s ability to produce different outcomes. This method is often referred to as a ballot box or vote counting method (Wolf 1986; Glass et al. 1981).

Second, standardized effect size scores were calculated for each study and for each outcome in it. The effect sizes were calculated from the studies by transforming the reported statistics (t values, F values, means, standard deviations, chi square values, etc.) into r values, using formulas developed by Hunter and Schmidt (1990). Further, an overall effect size score was created for each study by averaging the individual effect size scores and weighting by the number of individuals in the study (Schmitt and Klimoski 1991).

Effect sizes are correlations that test the null hypothesis of no treatment effect (Hunter and Schmidt 1990). The significance of the effect scores was assessed using 95% confidence intervals rather than using significance tests, following the suggestion of Hunter and Schmidt (1990). If the confidence
interval includes zero, the null hypothesis cannot be rejected, meaning the intervention had no effect on changing the outcome variable.

Studies reporting "no significant findings" without reporting the actual values of the statistics were coded as having an effect size score of .00, which was assumed to be the best estimate of the population effect size for studies reporting no effect. Although this conservative approach is not universally accepted, it is generally viewed as a way to avoid greater biases by omitting studies with nonsignificant findings (Wachter and Straf 1990). Glass et al. (1981) found that neither eliminating these effect scores nor altering the coding procedure changed the overall effect size by a significant amount.

Reliability of Coding

Ten of the 57 studies were randomly selected and recoded to produce an estimate of interrater reliability. The agreement reliability was 92% across all variables included in the meta-analysis coding.

RESULTS

Descriptive Findings

Of the 57 evaluations of alternative education programs included in the meta-analysis, 33% reported having racially mixed samples. The majority of the programs were alternative high schools (58%). Further, 61% of them targeted specific populations, either low academic achievers (42%) or delinquents (19%). The majority of alternative schools were in an urban school district (77%). Many of the studies did not provide detailed descriptions regarding daily length of the program and duration of the intervention. Of those that did, 84% lasted the full school day, and in 64% of the programs, the students had attended the alternative school for over one school year. (See Table 1 for a summary.)

Finally, seven studies reported both pre-post and comparison group assessments. This resulted in 64 unique research designs; 34 (53%) of these employed a pre-post design without a control or comparison group and 30 (47%) were either pre-post or post-only designs with a control or comparison group. Twelve of the 30 (40%) comparison group designs employed random assignment for program participation. The majority of the research designs used matching or convenience samples for comparison of program to non-program participants.
TABLE 1: Descriptive Findings

**Sample Characteristics**
- Programs reporting a racial heterogeneous sample: 19 (33%)
- Alternative high schools in sample: 33 (58%)
- Programs with specific target populations: 35 (61%)

**Intervention Characteristics**
- Schools located in an urban school district: 44 (77%)
- Programs with a daily length of a full school day: 27 (84%)
- Programs with a duration over one school year: 14 (63%)

**Methodological Characteristics**
- Studies having:
  - Comparison group research designs: 30 (47%)
  - Pre-post research designs: 34 (53%)
  - Random assignment to program participation: 12 (40%)

NOTE: Percentages are based upon the number of studies reporting these data.

**Effectiveness Results**

Three components were included in the examination of effectiveness results: the vote counting method, effect size scores, and the correlations of overall effect sizes with selected moderators.

**Vote counting method.** The vote counting method of review consisted of coding the reported results for each outcome as having a positive effect, no effect, or a negative effect. Table 2 presents the findings from this method, along with the number of distinct research designs assessing each outcome. The findings were grouped by the type of research design due to differences between studies employing a pre-post design and studies with a comparison group. These findings show that alternative education programs seldom produce negative effects. However, the distinction between positive effects and no effects appears to be dependent upon the research design. With the exception of school attitudes, the majority of the studies employing a pre-post research design found a positive effect, whereas the majority of studies with comparison group designs found no effects.

**Effect sizes.** The analysis of effect size scores presents a slightly different picture than the vote counting method (Table 3). It suggests that alternative education programs tend to have overall positive effects. The mean effect sizes for all of the outcomes across both research designs were greater than zero (with the exception of delinquency for both types of designs and school performance for comparison group designs). The largest weighted mean
**TABLE 2: Vote Counting Method Results**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>N</th>
<th>Positive Effect</th>
<th>No Effect</th>
<th>Negative Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>30</td>
<td>7 (23%)</td>
<td>20 (67%)</td>
<td>3 (10%)</td>
</tr>
<tr>
<td>Pre-post</td>
<td>34</td>
<td>23 (68%)</td>
<td>11 (32%)</td>
<td>0</td>
</tr>
<tr>
<td>Delinquency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>9</td>
<td>2 (22%)</td>
<td>6 (67%)</td>
<td>1 (11%)</td>
</tr>
<tr>
<td>Pre-post</td>
<td>4</td>
<td>2 (50%)</td>
<td>1 (25%)</td>
<td>1 (25%)</td>
</tr>
<tr>
<td>School performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>25</td>
<td>3 (12%)</td>
<td>16 (64%)</td>
<td>6 (24%)</td>
</tr>
<tr>
<td>Pre-post</td>
<td>31</td>
<td>19 (61%)</td>
<td>12 (39%)</td>
<td></td>
</tr>
<tr>
<td>School Attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>16</td>
<td>10 (63%)</td>
<td>5 (31%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Pre-post</td>
<td>6</td>
<td>3 (50%)</td>
<td>3 (50%)</td>
<td>0</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>13</td>
<td>4 (31%)</td>
<td>9 (69%)</td>
<td>0</td>
</tr>
<tr>
<td>Pre-post</td>
<td>10</td>
<td>7 (70%)</td>
<td>3 (30%)</td>
<td>0</td>
</tr>
</tbody>
</table>

effect sizes were found with pre-post research designs for school attitude (.49) and for the overall effect size (.33). The largest weighted mean effect size for comparison group research designs was school attitude (.20). With the exception of delinquency and school performance, all of the outcome variables across both types of designs fell in a 95% confidence interval that did not include zero. Findings with those confidence intervals can be interpreted as indicating that a nonzero effect would be expected between participation in the alternative school programs and school attitude, self-esteem, and the overall program effect.

Column F of Table 3 reports the percentage of observed variance due to sampling error. Hunter and Schmidt (1990) suggest that the percentage of observed variance due to sampling error is inversely related to the potential for finding moderator effects between the study characteristics and the effect size. Because the majority of the percentages are below 50%, a relatively high amount of variance across studies remained unexplained. It is possible that this remaining variance is due to other artifacts in the reviewed studies or some moderator variables (Schmitt and Klimoski 1991). To explore this possibility, correlations were analyzed between the overall effect sizes and selected program characteristics, including racial heterogeneity, grade level, target population, daily length of the program, and duration of the program (Table 4). Recall that these program characteristics (or potential moderator variables) were discussed in other literature reviews as being important in influencing outcomes of alternative school programs.
TABLE 3: Effect Size Results

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>30</td>
<td>6,986</td>
<td>.11</td>
<td>.08</td>
<td>.09</td>
<td>35%</td>
<td>.04 to .12</td>
</tr>
<tr>
<td>Pre-post</td>
<td>34</td>
<td>1,553</td>
<td>.36</td>
<td>.33</td>
<td>.26</td>
<td>21%</td>
<td>.23 to .43</td>
</tr>
<tr>
<td>Delinquency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>9</td>
<td>1,542</td>
<td>.00</td>
<td>.03</td>
<td>.09</td>
<td>41%</td>
<td>-.05 to .11</td>
</tr>
<tr>
<td>Pre-post</td>
<td>4</td>
<td>262</td>
<td>.40</td>
<td>.23</td>
<td>.42</td>
<td>7%</td>
<td>-.20 to .66</td>
</tr>
<tr>
<td>School Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>25</td>
<td>5,576</td>
<td>.06</td>
<td>.06</td>
<td>.12</td>
<td>24%</td>
<td>.00 to .11</td>
</tr>
<tr>
<td>Pre-post</td>
<td>31</td>
<td>1,515</td>
<td>.34</td>
<td>.27</td>
<td>.25</td>
<td>22%</td>
<td>.17 to .37</td>
</tr>
<tr>
<td>School Attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>16</td>
<td>2,457</td>
<td>.22</td>
<td>.20</td>
<td>.22</td>
<td>11%</td>
<td>.08 to .31</td>
</tr>
<tr>
<td>Pre-post</td>
<td>6</td>
<td>289</td>
<td>.38</td>
<td>.49</td>
<td>.15</td>
<td>36%</td>
<td>.31 to .62</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>13</td>
<td>3,067</td>
<td>.09</td>
<td>.07</td>
<td>.06</td>
<td>57%</td>
<td>.02 to .11</td>
</tr>
<tr>
<td>Pre-post</td>
<td>10</td>
<td>383</td>
<td>.34</td>
<td>.28</td>
<td>.17</td>
<td>44%</td>
<td>.14 to .42</td>
</tr>
</tbody>
</table>

NOTE: A = number of studies reporting this dependent variable; B = total number of subjects reflected in the effect size across all studies; C = the mean effect size; D = the weighted mean effect size; E = corrected variance of the effect size; F = percentage of observed variance due to sampling error; G = 95% confidence interval around the effect size.

TABLE 4: Correlations of Effect Size with Selected Study Characteristics

<table>
<thead>
<tr>
<th>Study Characteristic</th>
<th>Comparison Group Effect Size</th>
<th>Pre-Post Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racial heterogeneity</td>
<td>.10, n = 30, ns</td>
<td>-.14, n = 34, ns</td>
</tr>
<tr>
<td>Grade level</td>
<td>.08, n = 27, ns</td>
<td>.29, n = 33, .05</td>
</tr>
<tr>
<td>Target population</td>
<td>.41, n = 28, .01</td>
<td>.31, n = 33, .04</td>
</tr>
<tr>
<td>Program characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily length of program</td>
<td>-.03, n = 22, ns</td>
<td>.45, n = 16, .04</td>
</tr>
<tr>
<td>Duration of program</td>
<td>-.29, n = 11, ns</td>
<td>.18, n = 13, ns</td>
</tr>
</tbody>
</table>

NOTE: Coding for study characteristics: Racial heterogeneity: 1 = none, 2 = racially mixed sample; Grade level of students: 1 = elementary school, 2 = middle school, 3 = high school; Target population: 1 = none, 2 = yes; Daily length of program: 1 = after school, 2 = half a school day, 3 = full school day.

The only significant correlate of overall effect size across both research designs was whether the program involved a specific target population. This correlation indicated that alternative schools targeting a specific population...
(primarily low school achievers or delinquents) had significantly higher effect sizes than programs not targeting specific types of students. Furthermore, the overall effect sizes of pre-post research designs were positively correlated with alternative high schools (compared to alternative middle and elementary schools) and daily length of the program.

DISCUSSION AND CONCLUSIONS

This article used meta-analysis to assess the findings of alternative education programs in an effort to provide a more comprehensive summary of the literature than that found in prior reviews. The review consisted of a vote counting method and the calculation of effect sizes to determine the overall effect of the program as well as the programs’ ability to change delinquent behavior, school performance, school attitude, and self-esteem.

The results suggest that alternative education programs can have a small positive effect on school performance, school attitude, and self-esteem, regardless of research design. The highest pre-post and comparison group effect size was attitude toward school. This finding is consistent with prior research on alternative schools, which suggested that most students enjoy going to an alternative education program (Arnoke and Strout 1980; Ellison and Trickett 1978; Duke and Perry 1978).

The principal negative finding was that alternative schools have been unable to affect delinquent behavior. It is difficult to explain why this is so, given that school failure has been shown to be strongly related to future delinquency (Junger-Tas 1992). Based on the findings here, we speculate that even though alternative schools promote positive school attitudes, their effect on school performance and self-esteem is not large enough to influence delinquent behavior. That is, even though the students liked going to the alternative school and appear to have performed well, these gains did not overcome other influences that may have had a greater effect on subsequent delinquency (e.g., family and peers).

The relationships between the moderator variables and the overall effect sizes suggest that alternative schools with specific target populations have more impact than do undefined alternative schools. It is possible that these programs have developed a curriculum and structure around the needs of the target population, whereas other alternative schools end up with a mix of students with different problems and needs.

In addition, it appears that the research design played a major role in the conclusions reached by these studies. Both the vote counting method and the
effect size method show that pre-post research designs more often turn up positive results than do comparison group designs. In all likelihood, pre-post research designs are less rigorous and are more prone to internal validity threats (history, maturation, and statistical regression). Without a control or comparison group, it is not possible to determine how much of the effect was caused by the program and how much was attributable to other influences (Babbie 1992; Singleton, Straits, Straits, and McAllister 1988; Kerlinger 1986; Cook and Campbell 1979).

Methodological shortcomings continue to plague the alternative school literature. These center on the lack of true experimental research designs and the lack of extended follow-ups. It was the original intent of this review to examine follow-up studies; however, only three studies provided the necessary statistics to be included in the meta-analysis. This was particularly surprising, because what happens to alternative school graduates upon reentry into the regular school (or mainstreaming) has been widely cited as of primary interest (Frazier and Baenen 1988; Reilly and Reilly 1983; Barr et al. 1977).

This review was further hindered by poor reporting practices. First, many studies had to be omitted because the required statistics were not provided. Second, some studies in the meta-analysis did not give detailed descriptions of the programs. For example, only half of the 57 programs reported the intensity (daily number of hours at the alternative school) and fewer still mentioned the duration (months or school terms in the program). The analysis of moderator effects was also limited due to sketchy sample descriptions, which greatly inhibited the examination of the influence of program and sample attributes on effect sizes. Third, there are very few published studies regarding the effectiveness of alternative schools. The majority of studies included in the meta-analysis were found in ERIC as unpublished final reports. These studies contained many citations of other alternative school evaluations that could not be located. It is possible that the lack of published works is a result of the difficulty of publishing nonsignificant findings (Rosenthal 1979). Regardless, more emphasis needs to be placed on disseminating alternative education research.

The present review has demonstrated that alternative education programs can be beneficial to students in specific target populations, even though the poor condition of the literature has limited the ability of this review to fully investigate the influence of moderator variables on the effectiveness of these programs. Although this study has shown that alternative schools have had small positive effects on attitudes and performance outcomes, little is still known about why some programs are more successful than are others.
NOTES

1. This definition of an alternative school has been given by Raywid (1983) and Smith (1974).
2. The low number of empirical studies is not surprising. Other reviews have pointed out that the alternative education literature overwhelmingly focuses on the history and structure of alternative schools rather than on outcome variables (Garrison 1987; Barr et al. 1977).
3. Many of the citations, especially those found in ERIC, were final reports of several individual programs. For example, Gottfredson (1983) included the evaluation of several separate alternative education programs in one final report. In these cases, each program was coded as a separate study. This counting method produced 57 outcome studies from 27 reports.
4. Separate analyses were originally performed for self-reported delinquency and official delinquency outcomes. These measures did not produce different results and were combined.
5. Hunter and Schmidt (1990) suggest using confidence intervals over significance tests when interpreting effect size correlations, especially in reviews with small samples. Confidence intervals eliminate the necessity of establishing significance levels and are not as likely to be inflated by measurement error. This decreases the possibility of the investigator incorrectly interpreting the significance test.
6. Unfortunately, several studies did not report detailed demographic data. All that is known from this percentage is whether the study mentioned having minorities in the alternative school program. Studies did not indicate the percentage of the group that consisted of minorities or the ethnic group to which individuals belonged.
7. Eleven (19%) of the programs included middle school students and three (16%) were for elementary school students. Four (7%) programs did not report the grade level of the participants.
8. Thirty-two of the 57 programs reported daily length of the program, and 22 out of 57 reported duration of the intervention.
9. Delinquency was reverse coded for all of the analyses in order to maintain a consistent direction in the interpretation of program effects. Programs that are coded positive for delinquency should be interpreted as having a positive effect on reducing delinquent behaviors.

REFERENCES


